DÆDALUS

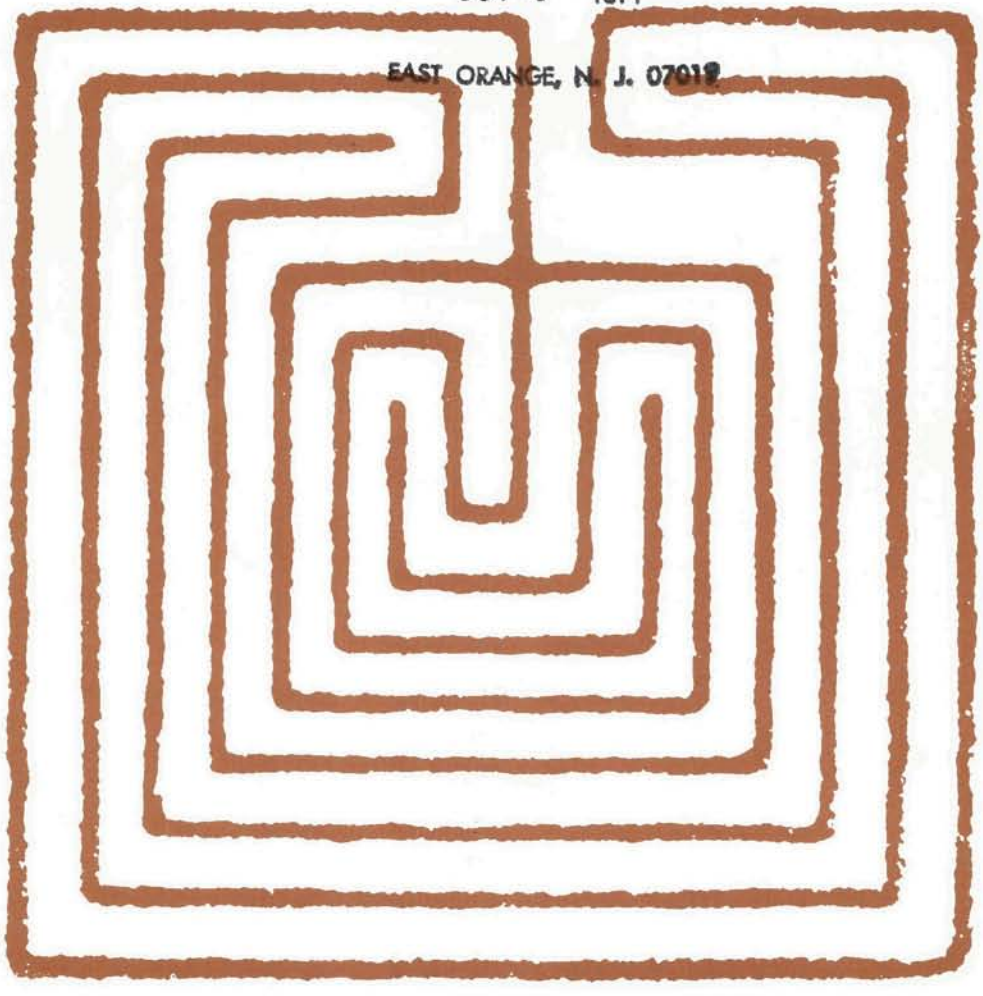
JOURNAL OF THE AMERICAN ACADEMY OF ARTS AND SCIENCES

Jonathan Culler
Shmuel N. Eisenstadt
Gerald Holton
Emmanuel B. LeRoy Ladurie
George A. Miller
Bruno Rossi
Jean Starobinski
Lester C. Thurow
Steven Weinberg
Edward O. Wilson

Discoveries and Interpretations
Studies in Contemporary Scholarship

Volume II

UPSALA COLLEGE
LIBRARY
OCT 4 - 1977
EAST ORANGE, N. J. 07019



Fall 1977

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

$3.50

Fall 1977: Discoveries and Interpretations: Studies in Contemporary Scholarship, Volume II. Issued as Vol. 106, No. 4 of the Proceedings of the American Academy of Arts and Sciences

Stephen R. Graubard, Editor of the Academy and of Daedalus; *Geno A. Ballotti, Managing Editor; Katharine O. Parker, Manuscripts Editor; Ruth Taylor, Subscription Fulfillment Manager*

Board of Editors

Daniel Bell, Isaiah Berlin, François Bourricaud, William Bouwsma, Asa Briggs, Herman Feshbach, Paul Freund, Clifford Geertz, Lillian Hellman, Gerald Holton, Peter L. Kapitza, Charles P. Kindleberger, Arnaldo Momigliano, Steven Muller, Gardner C. Quarton, Judith N. Shklar, Jean Starobinski

Editorial Office: Daedalus, *the Journal of the American Academy of Arts and Sciences, 7 Linden Street, Harvard University, Cambridge, Massachusetts 02138.*

Subscription Office: All subscriptions and single copy orders and change of address information should be addressed to the American Academy of Arts and Sciences, 165 Allandale Street, Jamaica Plain Station, Boston, Massachusetts 02130.

Subscriptions: Daedalus *is published quarterly. Subscription rate: $12.00 one year; $18.00 two years; $24.00 three years.*

Published by the American Academy of Arts and Sciences.

Printed in the United States of America. Printing Office: 2901 Byrdhill Road, Richmond, Virginia. Library of Congress Catalog Card Number 12-30299. Second-class postage paid at Boston, Massachusetts, and at additional mailing offices.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Discoveries and Interpretations

Studies in Contemporary Scholarship

Volume II

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

DÆDALUS
Journal of the American Academy of Arts and Sciences

Discoveries and Interpretations

Studies in Contemporary Scholarship

Volume II

GERALD HOLTON	v	*Introduction*
JEAN STAROBINSKI	1	*Criticism and Authority*
STEVEN WEINBERG	17	*The Search for Unity: Notes for a History of Quantum Field Theory*
BRUNO ROSSI	37	*X-Ray Astronomy*
SHMUEL N. EISENSTADT	59	*Sociological Theory and an Analysis of the Dynamics of Civilizations and of Revolutions*
LESTER C. THUROW	79	*Economics 1977*
JONATHAN CULLER	95	*In Pursuit of Signs*
GEORGE A. MILLER	113	*Problems of Communication*

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

EDWARD O. WILSON 127 *Biology and the Social Sciences*

EMMANUEL B. LEROY LADURIE 141 *Recent Historical "Discoveries"*

FALL 1977: DISCOVERIES AND INTERPRETATIONS: STUDIES IN CONTEMPORARY SCHOLARSHIP, VOLUME II

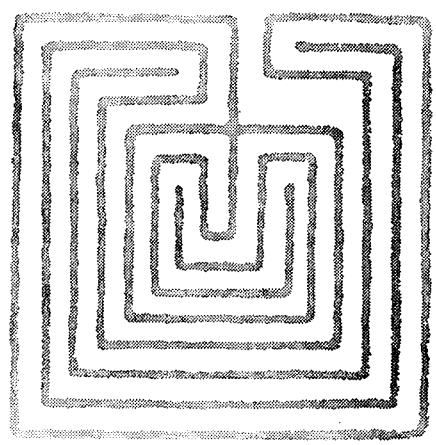
Issued as Vol. 106, No. 4 of the Proceedings of the American Academy of Arts and Sciences

This Special Issue was prepared under financial support of the National Science Foundation in connection with its Bicentennial "Knowledge 2000" project. Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of NSF.

All Subject Data shall be in the public domain. None of the material produced under this contract shall be copyrighted.

Thanks are due also to the Ford Foundation and the Rockefeller Foundation, whose generous support helped make this Special Issue possible.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED



PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

GERALD HOLTON

Introduction

ABOUT TWO YEARS AGO, the editor of *Daedalus*, Stephen R. Graubard, asked me to join him in the first stages of planning the summer and fall 1977 issues intended to review some of the achievements in contemporary scholarship. As we were thinking through the great spectrum of candidates—of research fields from which all too few had to be chosen, of persons whom to involve in further planning and as authors—we were struck by the realization that these issues would mark the twentieth year of publication of the quarterly.

Can it really be all of twenty years? Dwelling on the personal meaning such an anniversary has for all who have been working to produce the 80 issues of the journal might nowadays invite the charge of sentimentality. Still, I confess that my thoughts returned to the very first quarterly issue, sent to press in late summer 1957. Although we did not know it at the time, we were in that brief golden age for scholars, between the downfall of McCarthyite anti-intellectualism and the failure of support and trust during the later phases of the Vietnam war. In fact, we were in the state of pre-Space (and indeed pre-Snow) innocence; yet, the topic of that first issue was nothing less than *Science and the Modern World View*.

The contributors were humanists, natural scientists, and social scientists who had met to present their papers in homage to the work of two distinguished colleagues, the scientist-philosophers P. W. Bridgman and Philipp Frank. It is of some interest to look back, to see what was then of chief concern to such a group of scholars, and to compare it with the new preoccupations emerging from the essays in the current issue, two decades later.

Because of their common interest in the "unity of science" movement, the chief question before the authors in 1957 was what hope there might be in the eventual achievement of some common understanding across the disciplines, to take the place of the unified world view that once existed, or at least seemed to have existed. What, Frank asked, could our discourse on politics or ethics learn from the successes that contemporary science has had in the "techniques for analyzing the meaning of terms"? Conversely, if one agreed that science consists not in producing abstractions from experience but in the invention of symbols and the building of a symbolic system—a "work of creative imagination which acts on the basis of experience"—should one not expect in principle a fundamentally similar view of "reality" among scientists and artists?

The ecumenical goal seemed attractive to all authors, although Bridgman expressed his pessimism with respect to hopes for any early, substantial accomplishments along that line. Recent scientific experience had helped show to some degree "what is involved in the proper use of our minds." It has exposed a whole list of "illegitimate concepts," such as "certainty." But outside science,

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

the situation was that of a system trying to deal with itself, and that seemed to him contrary to the "first law of mental dynamics," namely, "the law that we cannot get away from ourselves." In science and out, "the human race has not yet found how to use its mind. . . . The sciences and the humanities find themselves facing the problem together." At best, he concluded, we may have by now at least "advanced to the point where we can put our hand on the hem of the curtain that separates us from an understanding of the nature of our minds. Is it conceivable that we will withdraw our hand and turn back through discouragement and lack of vision?"

J. Robert Oppenheimer addressed himself at the meeting of the authors head-on to

> one way of bringing a wider unity to our time: That unity, I think, can be based only on a rather different kind of structure than the one that most of us have in mind when we talk of the unity of culture. I think that it cannot be an architectonic unity, in which there is a central chamber into which all else leads, the central chamber which is the repository of the common knowledge of the world. . . .
>
> I think that the unity we can seek lies really in two things. One is that the knowledge which comes to us at such a terrifyingly, inhumanly rapid rate, has some order in it. We are allowed to forget a great deal, as well as to learn. This order is never adequate. The mass of un-understood things, which cannot be summarized, or wholly ordered, always grows greater; but a great deal does get understood.
>
> The second is simply this: we can have each other to dinner. We ourselves, and with each other by our converse, can create not an architecture of global scope, but an immense, intricate network of intimacy, illumination, and understanding.

With his usual eloquence, Oppenheimer had pronounced a kind of farewell to the dream of the unity of science movement. That movement—more properly, the "unified knowledge" movement, as indicated by its original designation, *Einheitswissenschaft*—had a long history. One of its many projects, an *Encyclopedia of Unified Science* that would systematize and connect the fields of learning, had been planned at least since 1920 in discussions among Otto Neurath, Einstein, Frank, Hans Hahn and Rudolf Carnap. The first two volumes (out of a projected 36) were now nearly completed, with published contributions by Niels Bohr, John Dewey, Bertrand Russell, and many other distinguished scholars.

One now looks back on these plans with admiration and envy. But Oppenheimer was right; today it is obvious what then seemed still worth struggling against: that a synthesis on so large a scale is possible neither within the individual person nor among the several fields of learning. As a consequence, however, the need is greater than ever to recognize how small one's own portion of the world is, to view from one's own narrow platform the search of others with interest and sympathy, and so attempt to re-establish a community of learning on the recognition that what binds us together is perhaps chiefly the integrity of our individual concerns.

"Unity" Ended?

The "unity of knowledge" is now a subject of interest to a few historically minded philosophers rather than an explicit research goal shared by specialists

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

in various fields. Owing to the cyclical nature of such interests, more attention will no doubt be given to another noble effort again, eventually. For the time being, however, one finds only superficial obeisance to the idea; another conference on this topic, with major participants like those twenty years ago, is hard to imagine.

Perhaps for the same reasons, one can count on the fingers of one hand those among our foremost scientists and scholars who have recently written books that seriously make the experience of modern science and scholarship—the act of doing the work as well as something of the spirit motivating its practitioners—available to a wide readership. Volumes of this sort have not necessarily or instantly produced universal enlightenment in the past. But one must ask oneself why now, perhaps for the first time in a century, so little attention is given to producing this kind of work.

There is a paradox here: it starts with the fact that science today—and for the moment I shall restrict myself to science—is without doubt more powerful, more interesting, indeed more exhilarating than it has ever been, both as a product and as a process. As the essays in these very issues of *Daedalus* show, the intellectual constructs are more far-reaching in their control over the phenomena, and the techniques are more sophisticated than have been thought possible even recently. A relatively few themata and metaphors provide the simple armatures that hold up seemingly complex structures in widely different specialties—somewhat as the elementary concepts of plate tectonics account for the main features of both the Rocky Mountains in the north and the Andes in the south.

One would therefore expect that the commonalities of the sciences, in particular their shared philosophical underpinnings, would now be more evident than ever, and the claim of the sciences as natural elements in the total cultural ecology more easily granted. But that is precisely what we do not find. At the very moment when basic research in a great variety of fields is at its best by its own criteria, the whole enterprise is subject to doubt about its larger aims more than at any time in the recent past.

Where is the whole tradition of philosophically sophisticated scientists who shared their views on the powers, limits, and "Natural-Philosophy" element of the enterprise, readably and without condescension, addressing themselves to a wide public? One thinks here of Poincaré, Mach, Hadamard, Einstein, Bohr, Schrödinger, Jeans, Eddington, Bridgman, Frank, Oppenheimer, to name just a few in the physical sciences. Until a generation ago, no one who claimed to be intellectually civilized would have dreamed of neglecting or turning his back on what these men were saying. They, in turn, saw it as their business to give a humanistically informed account of the claims, products, and processes of their work. Today, only a few of the best can persuade themselves to write even a generally readable account of their own technical specialty for the popular science journals, those valiant umbrellas against the rain of knowledge.

As a result, the overwhelming majority of students and laymen, even the intellectually most alert, has read little by scientists except school texts. I do not believe that members of the general public have lost the ancient interest in works with unifying intent. On the contrary, the desire to understand the place and thrust of science leads them to the kind of substitute literature that has filled

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

the vacuum—a variety of views at second hand, sometimes verging on science fiction, and the popular arguments condemning science and technology, presented often with considerable eloquence.

All this is not accidental. The demise of a whole genre of essential works that so many of us read in the past has identifiable causes. One of these, surely, is a change in role perception. For better or worse, the premier scholar thinks of himself no longer as a *Kulturträger* with all the civilizing functions and obligations that this role implied. It is difficult enough to deal adequately with the complex demands of the day-to-day job. The bright prospects that beckon in his study or laboratory pull him in one direction, the murky and clamorous issues of science and public policy or science and society try to engage him on the other side. At the meetings of professional societies, time left over from specialty discussions has to be given to practical concerns—about the threat that the combination of insufficient funding and red tape is causing the high quality of research to be undermined; that the careers of young persons whom we are training are threatened by such a narrowing down of opportunities in academic life that a "lockout" of younger people is predicted for the next few years; that educational innovation all too often becomes the victim of purely bureaucratic rearrangement or cost-cutting. Against such clear and present realities, the needs of nurturing the unifying connections between the sciences, social studies, and humanities is a luxury that is constantly losing in priority. As the president of the National Academy of Sciences recently said, "I believe there is a greater consciousness of this gap, but it widens, if anything. Where once there was an attempt to build bridges, one now finds distrust and alienation."

It is in fact the public rather than the scholarly community which now forces attention on such issues as the need for a clearer understanding of scientific activity and its role in society. During the recent debate in the Cambridge, Massachusetts City Council on the construction of a laboratory for research on recombinant DNA, there were the usual occasions for grilling the scientists who had come to testify. ("Can you make an absolute, 100 percent certain guarantee that there is no possible risk which might arise from this experimentation? Is there zero risk of danger? Answer that question later, please. . . . Is it true that in the history of science mistakes have been made? Do scientists ever exercise poor judgment? Do they ever have accidents? Do you possess enough foresight and wisdom to decide which direction the future of mankind should take?") But the most poignant comment came from a city councillor who reflected on how poorly he and his constituents were prepared to deal with scientific issues. He quietly turned to the scientists there with the remark that they had not only done little to prepare their audience to think of the dangers that certain scientific research may or may not have—they had not even bothered to make plain and interesting what good things the scientists were doing. He was making the essential point that with the political process now entering into the life of the academy more and more prominently, it is doubly necessary that the scientists and scholars be very clear about their work, their intellectual ambitions, the process of discovery, and its needs, powers, and limits.

It is a reasonable hypothesis that the disappearance of the genre of which I spoke, coming at just the time when it may be most needed, is in good part caused by the increasing complexity and scale of scientific and other scholar-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

ship, and of the relation of scholarship to society in our time. Precisely the most thoughtful scientist is likely to see that the proper accounting of science, where it stands in contemporary thought, and how it may best fulfill its aims, requires a professional perspective that many working scientists do not have, that in fact simply was not available until fairly recently. That is, an ambition of the sort which motivated a Poincaré requires one now to have some commands of the working tools and main literature not only of one's own science but also of aspects of the history and sociology of science, subjects that have become thoroughly professionalized precisely since the days of the "unity of science" movement.

Functional Unity by Incorporation

The head-on approach of the movement evidently did not work, or at least is slumbering for another day. The tools of the logical empiricists—foremost among them the scissors for paring away nonfunctional terms and concepts in order to reach an *Einheitswissenschaft*—did much, but then became dull. For another self-conscious attempt at some large-scale synthesis, we shall have to wait a very long time. But the case is by no means hopeless. Almost unremarked, more modest unifying processes are at work, more effective than they have ever been, both within specialty fields and across the borders of such fields.

To me, that is a chief lesson of the essays before us here and in the previous issue of *Daedalus*. These essays were not commissioned with a view to exhibiting unities. On the contrary, the Editor suggested to each author to take stock of what is today most vital or problematic in his field of scholarly research, and to attend both to the limits to recent progress and to the conceptual advances, without attempting a complete survey. Yet, among the many other ways to read and react to this garland of presentations, and quite apart from reading them as reports from the frontiers of intellectual excitement in the separate fields chosen here, virtually all of the essays indicate the complementary activity of both generalization and specialization. The latter represents the drive toward individualization and particularization, often lamented and yet the life blood of scholarly advance. The former is a drive to synthesis rather different from that of *Einheitswissenschaft*, but possibly just for that reason more effective today. I would call it *functional incorporation*. It is a form of synthesis within fields, and sometimes among specific aspects of different fields, which stems from the fact that so much of modern scholarship is problem oriented rather than discipline oriented. The mandate for incorporation, for consolidation and interpenetration, is thus primarily functional rather than structural, for it comes from the here-and-now requirements of solving a problem, rather than from an overarching and ideological commitment to reorganize the bases of all scholarship.

A glance at each of the nine essays in succession from this point of view will indicate the operation of this functional drive. Thus, Jean Starobinski, writing on criticism, notes that in the hands of the critic willing to do so, philology can, even if renouncing its precise and limited problems, "by reabsorbing itself into the 'human sciences' [sociology, psychology, linguistics], by almost melting in-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

to them," yield profits "under the column of general theories of culture, literature, etc." He cites Ernst Renan, who regards it as the "new dogmatism" that historical scholarship "is one of the aspects of the progress of reason, at the end of which humanity, conciliated and divinized, will enter into the great unitary realm of achieved Science."

As Plato had already urged (e.g., in *The Republic*, 509–511, 533–534), the critic-scholar is admonished not to rest content with either analysis or synthesis alone, but to use both in proper order: "Although the totality of meaning, that asymptotic object of completed knowledge, might seem the opposite of the minute fact isolated by the philologist, it is nevertheless here, in the fine detail which the scholar discerns and adds to other analogous discoveries, that the former kind of knowledge has, according to Renan, its starting point. It is from this fragmentary starting point that inductive work commences its course through the universality of the 'sciences of man'. The final 'diamond' remains, of course, a hope that requires an act of faith, like the promise of Paradise: history will culminate for everyone, in intelligent plenitude."

No field, on the surface of it, seems further removed from Starobinski's than that with which Steven Weinberg deals in "The Search for Unity: Notes for a History of Quantum Field Theory." Yet, the very title supports the main point I am trying to make. Here, we have a history of the spectacular effort to find the basis underlying the chief fields of physical science, the search for a general framework of quantum field theory into which one can incorporate all four forces of nature—the weak, electromagnetic, and strong interactions, and perhaps even the recalcitrant gravitational one. This is not merely a matter of analogy; the theory tries to unify the fields by treating them as members of a single family.

The main tool and guiding principle which Weinberg employs to "pick the quantum field theory of the real world out of the infinite variety of conceivable quantum field theories" is to impose a requirement of great restrictiveness (here, the requirement of renormalizability). The chief thema at work is that of *necessity*. One remembers that the grand theories of physical science were chiefly persuasive insofar as the proposals exhibited just this kind of restrictiveness. Thus there was nothing arbitrary, no room for the smallest ad hoc rearrangement of the orbits in the Copernican solar system; it revealed a sparse rationale, a necessity that binds each detail to the whole design. Hence it carries the conviction that we understand why the planets are disposed as they are and not otherwise.

Weinberg gives us a glance at the quiet underground current of the same cosmological tradition, although it now comes in somewhat disguised form. Elsewhere, he has elaborated eloquently on his motivation for a search for unity through the application of a drastically restrictive principle. On accepting the J. Robert Oppenheimer Memorial Prize recently, he said:

> Different physicists have different motivations, and I can speak with certainty only about my own. . . . The reason I take such an optimistic view of where we are now is that relativity and quantum mechanics, taken together but without any additional assumptions, are extraordinarily restrictive principles. Quantum mechanics without relativity would allow us to conceive of a great many possible physical systems. Open any textbook on nonrelativistic quantum mechanics and you will find a rich variety of made-up examples—particles in rigid boxes, parti-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

> cles on springs, and so on—which do not exist in the real world but which are perfectly consistent with the principles of quantum mechanics. However, when you put quantum mechanics together with relativity, you find that it is nearly impossible to conceive of any possible physical systems at all. Nature somehow manages to be both relativistic and quantum-mechanical; but those two requirements restrict it so much that it has only a limited choice of how to be.

The attractiveness of a system with such qualities has been understood since the beginning of science. And although any individual attempt of this sort is an act of intellectual and professional risk-taking—for the thematic choice itself is neither verifiable nor falsifiable, and the antithetical thema of complexity, for example, has not been conquered thereby—no other, less daring approach is likely to lead to a synthesis, least of all to one having the sweep that historically has given the scientific enterprise its intellectual mandate.

In Bruno Rossi's article on X-ray astronomy, one may observe the mechanism of coherence again, now working among neighboring fields of physical science—physics and astronomy. In addition, we have here a much needed reminder that the same integrating forces exist between engineering (in the production of ingenious measuring devices) and science. It would be difficult to decide in which of these fields this work is most interesting; at any rate, the achievement in any one of them would have been impossible without the simultaneous achievement in the others.

Shmuel Eisenstadt, from the very outset, marks off for himself as encompassing a field as can readily be imagined: the study of societies and civilizations. His analysis of the dynamics of patterns of change stresses the "cultural and symbolic dimensions and the sociohistorical one," and he regrets that only for reasons of space he could not also consider ecological factors. The sweep of application of his approach is correspondingly large. It can be applied in principle "to any field of sociological analysis, ranging from macrosocietal studies to those of organizations and small groups, to patterns of communication."

Lester C. Thurow's essay on "Economics 1977" again provides impressive evidence of the integrative drive at work in his field. For, despite the dour view one may take of it (as he does), economics surely does not lack outreaching ambitions. There is, he points out, an "economics of crime, marriage, suicide, and other social phenomena that were once regarded as the exclusive preserves of other social sciences." Moreover, "economists have had their imperialism accepted by much of the rest of the world," and have now infiltrated law and medical schools, and departments of social work, engineering, urban and architectural studies, to name only a few.

Jonathan Culler, in his essay on the advent of contemporary semiotics, claims an even larger domain, the disciplines of *les sciences humaines* within which it is futile to distinguish the humanities from the social sciences. It is not surprising that a field which counts among its parents Pearce, Saussure, Whitehead, Langer, and Lévi-Strauss, has a vast outreach and a correspondingly powerful urge to incorporation. As Culler writes:

> If the "dissolution" of man into a series of systems is the ultimate effect of a structuralist and semiotic perspective, if everything which is meaningful within human

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

> cultures can be treated as a sign, then as [Jean-Marie] Benoist's list already suggests, semiotics embraces a vast domain: it moves in, imperialistically, on the territory of most disciplines of the humanities and social sciences. Any sphere of human activity, from music to cooking to politics, can be an object of semiotic study; and it is precisely because any signifying activity calls for semiotic investigation that the emergence of semiotics may bring about a major reorganization of scholarly research. If the study of music as a sign system is assimilated to semiotics, the other aspects of musicology will form a discipline which must define itself in a new way.

George A. Miller, on "Problems of Communication," begins with a remark that "in the imaginary matrix of problems-by-disciplines, the role devoted to it [communications] would have contributions in almost every column." He wonders: "How did we come to see so many different things as instances of communication?" The preoccupation is equally prominent in computer science, the social and behavioral sciences, linguistics, sociology and political science, psychology, anthropology, neurology, zoology, and philosophy. And he adds very justly: "In an age of analysis, therefore, the study of communication has provided a special refuge for synthesists, for those personalities whose response to disciplinary boundaries is a feeling of intellectual claustrophobia." The special advances he has chosen to discuss appeal to him precisely because he has always "enjoyed transporting ideas across provincial lines."

The separation between one parochial field and the next at an adjacent level of organization is at the very center of E. O. Wilson's concern in his essay, "Biology and the Social Sciences." Going far beyond the remarkable synthesis of Mendelian genetics and biochemistry which has been so recently accomplished in biology, and going even beyond his own well-known work on the integration of fields that has been given the name sociobiology, he now envisages a "juncture" of neurobiology and sociobiology with social science. He focuses on the creative tension between neighboring fields whose relationship makes them act as "antidisciplines." It is a challenging notion, of which a good deal more will be heard. He defines it in this manner: "By today's standards, a broad scholar can be defined as one who is a student of three subjects: his discipline, the lower antidiscipline, and the subject to which his specialty stands as antidiscipline [at the next level of organization]. A well-rounded neurophysiologist, for example, is deeply involved in the microstructure and behavior of single cells, but he also understands the molecular basis of electrical and chemical transmission, and he hopes to explain enough of neuron systems to help account for the more elementary patterns of animal behavior."

Historically, Wilson sees biology and chemistry as having passed through a period in which they were antidisciplines, now synthesized in certain areas into biochemistry. Biochemistry in turn has explained much of the cellular mechanisms that puzzled early cytologists, and so was transformed in part into molecular biology. During this period of the professional field's evolution, "progress over a large part of biology was fueled by a competition among the various attitudes and themata derived from biology and chemistry—the discipline and its antidiscipline." He adds a series of examples of what the "reverse side of the social sciences" looks like, "as it has been glimpsed by a biologist approaching in that direction": biological anthropology, where the central question is the na-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

ture and strength of the coupling between cultural and biological evolution; psychology and psychoanalytic theory, within which the logical role of sociobiology is to "reconstruct the evolutionary history of the structure of the unconscious"; economics, understood from so general a perspective that the conventional treatment of the subject is merely "the description of economic behavior in one mammalian species with a limited range of the biological state variables."

E. B. Le Roy Ladurie devotes his essay to "Recent Historical 'Discoveries.' " Like all the other essays, this too is meant to be, and should be, read primarily not as a case example of the present-day example of the mechanism of incorporation of subspecialties. And yet, that is again a strong theme running through this essay, as through all the others. In fact, in Le Roy Ladurie's own words, "Unlike the exact sciences and other human sciences, history rests at the very center of an extremely crowded, multiple crossing." Thus, discussing the accomplishments in social history of the medievalist scholar, Philippe Contamine, Le Roy Ladurie shows that Contamine's work is as much in ethnography, anthropology, and the sociology of professions as in history proper.

Analysis and Synthesis as Methodological Themata

A very similar case can be made about the essays in the previous volume, if additional evidence were needed. But it may now be useful to ask what such cases tell us about the current state of that eternal promise of a meaningful weaving together of different fields. It would be well to distinguish between four general meanings of the term: synthesis used in the *reconstitutional* sense (e.g., where an analysis, followed by a synthesis, reestablishes the original condition); synthesis used in the *judgmental* sense (as in the Kantian categories and their modern critiques); synthesis used in the *transformational* sense (e.g., where the application of analysis and synthesis advances one to a qualitatively new level); and, finally, synthesis used in the grand, *cultural* sense.

It is the last of these that seems to have been highest on the hidden agenda of the "unity of science" group and that now seems more elusive than ever. In this meaning of the term, we are talking about the kind of work which has traditionally defined the best in our culture, the synoptic product that characterizes the thinking of a period or a field. One thinks of the treatises of Aristotle and Aquinas, Spinoza and Kant, or the major literary works that have the same unifying character and intention, from the Greek epics to Goethe and Tolstoi.

With all their differences, these cultural products provide a framework of interpretation of particulars that help to propel thought and feeling to important truth. But we should note two curious asymmetries. Synthetic works as high cultural achievements are supposed to pervade the training and the consciousness of educated persons from an early age, whereas works of explicit analysis rarely have that high status. But when we turn to examine the conception from the point of view of praxis, as expressed in professional, scientific, and scholarly activity, or in the organization of departments in universities or funding agencies or of classes of Academy membership, the positions of analysis and synthesis are entirely reversed, and it is analysis which has the more prominent

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

place. A working intellectual will now only rarely claim to be concerned with "synthesis." Perhaps in view of their praxis-oriented purposes, even the major encyclopedias tend to have elaborate discussions of analysis but little on synthesis (neither the *Encyclopedia Britannica*, subtitled "Knowledge in Depth," nor the *Dictionary of the History of Ideas* has even an entry for this term).

Possibly the asymmetrical evaluation of analysis and synthesis is the result of the preponderance and success of reductionistic thought in our time. It may also reflect the discouragement with the achievement of any coherent statement emerging from the immensely complex patchwork of highly developed competences and antithetical interests.

Living without a realistic hope for a widely accepted and shared cultural synthesis entails, no doubt, some psychic costs. It may well be responsible for the sense of loss of the intellectual's role that once existed or at least was thought to have existed. But the alternative is not so bleak, if I am correct in this interpretation of shared key features of scholarship, reflected in these essays, and of the fruitful operation of integrative forces, at least among neighboring fields. The process of incorporation among disciplines, antidisciplines, and subordinate specialties has at least the character of a transformational synthesis.

Looking ahead along the trajectory traced out by the progress of contemporary scholarship toward its likely state another two decades hence, I for one have little doubt that the integrative forces will have asserted themselves even more. To quote Oppenheimer again, we "can create, not an architecture of global scope, but an immense, intricate network of intimancy, illumination, and understanding." At that hypothetical, next celebration of the state of scholarship and research, everything will still not be connected with everything else in the world we live in; but the network will undoubtedly be richer and stronger.

One recalls here a story told by the architect LeCorbusier. He had invented his "Modulor," a measure system for fixing the dimensions of architectural space, of urban design, and of plastic arts, by which all these would attain an overarching common form. Believing with fervor in the necessity and power of his synthesis, he was arguing everywhere for its wide adoption. He even journeyed to Princeton to enroll Einstein. Einstein listened patiently; but instead of the hoped-for endorsement, he made a milder and more realistic prediction, one that applies here also. The scheme, Einstein said, will be quite satisfactory if it merely serves to make the good easier, and the bad more difficult.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

JEAN STAROBINSKI

Criticism and Authority

I

ONE OF THE BEST ways to understand the role and the function of criticism in our time is by means of the history of the diverse meanings of the term *criticism*.[1] This understanding, based on semantic and conceptual history, will not oblige contemporary criticism to remain today what it once was: history is not cumulative, and the different meanings that one word may have accumulated over successive periods of its life show that some functions have been abandoned in favor of new purposes. Whatever their survival, the older meanings of this word are bound to bear the superimposition of a new meaning, to become its more or less distinct foundation. On the other hand, whatever the vividness of this new meaning, it instantly manifests itself by contrast with the previous sense which it supersedes and finally supplants. In semantic history, therefore, there is no such thing as total conservation or genesis ex nihilo.

At times older meanings are simply set aside; at others, they wither away completely. For the historian, each of these accidents bears signification, and helps clarify the specific value of the current meaning.

Among the many meanings of the word *criticism* that have now become quite blurred, there is one that, over several centuries, has corresponded to the activity of *restoration*.[2] Surely everybody knows what a "critical edition" is, but just as surely, nobody immediately thinks of this kind of activity when confronted with the term *criticism*.

And yet, in the eighteenth century the story was quite different. This is the age—the last of the so-called "classical centuries"—that I shall take as the starting point for my study.

In his theoretical writings (consisting, for the most part, of his contributions to the *Encyclopédie*) the otherwise mediocre writer Marmontel can claim credit for having clearly defined the concepts of usage current among his contemporaries. Nothing is clearer than his formulation of the "two general points of view" from which criticism can be approached:

> . . . One is the kind of study to which we owe the restoration of ancient literature. To have an idea of the importance of this work, it is enough to picture the chaos in which early commentators found the most precious works of antiquity. . . . The

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

> other point of view is to consider criticism as an enlightened analysis and an equitable judgment of human productions. All human productions can be gathered under three main rubrics: sciences, liberal arts, and mechanical arts. . . .[3]

It is quite evident that the second point of view distinguished by Marmontel is criticism in the broad-reaching and all-embracing sense that it will acquire for us: as a rule, no activity, no human "production" can escape its analysis; in Marmontel's time, criticism (taken in this second sense) was already including political and social institutions in its scope, although it was preferable not to speak of these matters too loudly. The double definition proposed by Marmontel distinguishes between a restricted meaning—philological, limited to the establishment of ancient texts—and a broader meaning encompassing all those acts of judgment applied to the analysis of the merit, the rationality, and the legitimacy of "human productions." Initially, one might well consider the first meaning—restricted, philological—to be a very particular case, an almost professional specialization of the exercise of judgment, that is, of criticism in its widest sense. But in fact, the historian will soon realize that it is this restricted meaning which, in the semantic evolution of the word, precedes and produces by extension, by some sort of contagious generalization, that broader meaning of which it is, as it were, the underground wellspring. As is often the case, the success of the derivative meaning undermines the original meaning until it quite obliterates the filiation that connects them. That the restricted philological meaning of the word *criticism* was already waning in the eighteenth century (by which I mean that it referred to an activity that was already considered in decline, and hence, that it became less important than the meaning which applied to activities in expansion) is clearly evidenced by Marmontel's observations:

> The restorers of ancient literature had only one object, and an uncertain one at that; it was to render the authors mutually intelligible with the help of monuments. But in order to pass this ancient gold on to us, many had to die in the mines. Let us admit it, we treat this kind of criticism with much too much contempt, and those who have pursued it so laboriously for themselves and so usefully for us, with much too much ingratitude. Enriched by their vigilance, we glory in the possession of what we would want them to have acquired ingloriously. It is true that, since the merit of a profession depends on its combined utility and difficulty, scholarship has lost some of its prestige as it has become easier and less important; however, it would be unjust to judge what it once was by what it now is. . . . This part of criticism would also involve the verification of chronological calculations, if such calculations could be verified, but the poor results obtained by illustrious scholars in this field prove that to return to their research henceforth would be both useless and laborious. One must accept ignorance of what one cannot know, and it is very plausible that what remains unknown about the history of the ages will never be known, and the human mind will not be worse off for it.[4]

According to Marmontel, the task of restorative criticism has been achieved—unless (as in the case of chronological research) it is unachievable. If he speaks of it at the beginning of his article, it is at once to pay it his respects and to take his leave.

Let us try to show what restorative criticism essentially was until the day Marmontel (and many others along with him—d'Alembert, Voltaire, Diderot) thought it time to affirm that it had forever left its heroic age behind.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

The "restoration of the classics" is "the renaissance of letters"—what we call the Renaissance. And what scholarly humanism wants to restore are those texts to which sovereign authority is attributed: the Bible, Plato, Hippocrates, the Church Fathers, etc.—texts that must be recovered in their purity, before they were altered, corrupted, deformed, and apocryphally augmented. Given the superior degree of truth attributed to the authentic text such as it was dictated by the Holy Spirit or conceived by its original author, literary exactitude is absolutely vital. It is obviously the sacralization of the origin and of the original that impels the rejection of any superadded word as some malefice or vice; interpolations, copyists' errors, and questionable translations are all obstacles to be removed. To go to the source will be to receive the original enlightenment of truth directly, without the interference of any recent (and hence usurpatory) authority. It has often been observed that, through its direct return to the original texts of the great books, humanism challenges the tutelage and the restrictive interpretation of the Roman Church. Humanism attacks the interposed word in order to attain the initial one. Criticism, to use Marmontel's image, resembles the underground search for gold, unless one prefers to compare it to the crucible where gold is rid of all its dross.

Thus a direct relationship between conscience and the restored text becomes possible. Reading becomes a crucial encounter. This important point needs to be emphasized: the patient labors of the philologists as well as the fastidious task of collating texts, at last find their justification in the direct relationship they facilitate. Henceforth reading can be confident; it can assume the characteristics of faith without risking to go astray: for it is through reading that the soul of the believer will confront God's commandments, speak with him, and wholly adhere to him. The arduous task of returning to the source is thus legitimized by the unsurmountable value attributed to its original manifestation.

Emphasis is here laid exclusively on inner experience, for it is there that everyone feels that the Word addresses him in his unique destiny. And soon the attention of the reader will turn from this discourse of the Book which reading interiorizes to his own response, to the quality (pure or calculated) of his spiritual act. If, during his submission to the text, the reader hears an inner voice that acquiesces to the text and gives it its full agreement, he will soon believe that an identical law speaks within himself. The solitary reading of the sacred text marks the beginning of a transfer of authority which will later consecrate the dictates of the lone inner voice, or, to speak in Rousseauistic terms, the reading of those characters which Nature (or God) has engraved in the heart of man. In order to get to this point, it will be necessary to reactivate the critical process which, this time, will be directed entirely against all exterior law, even if it be the one ordained by the revealed Book.

At this juncture, the text, no matter how authentic, itself becomes suspect. For, however admirable, it remains a human work that has not only passed through several hands before reaching its reader, but one that was already deformed in its original state.[5] The only true text, the sole authenticity, is the self [*le moi*] listening to its heart: the individual is thus led to believe that the reading of the Book was not a genuine immediate revelation and that alone the voice of feeling speaks in him without distance. It is to himself, to his true being, that he now longs to be restored. And whereas the reestablishment of the text entailed discarding all inveterate errors, defective readings, interposed glosses, now it is

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

more advisable to discard any exterior word, all previous judgments that culture and education may have accumulated.

Criticism is thus no longer philological; its aim is no longer that of establishing the words and phrases of an earlier Book in their unshakeable objectivity. Now it turns against everything that does not bear the seal of a self-confident subjectivity, against everything that is not consonant with the voice of the heart or that of pure reason. And just as the Holy Book demanded paraphrase and commentary at a respectful distance, so the self that is heir to the Book now demands a literature subordinated to itself, that praises its "depths," while declaring them ineffable. In fact, the whole of literature becomes this expression and this paraphrase, destined to further commentary in the secondary, derivative literature of modern criticism, in the "literary portrait," the "analysis," or in "psychological criticism."

And yet, the critical tool employed by the Renaissance possesses an applicability, a power of action that in itself has no reason to stop at texts endowed with sacred prestige. From the Bible to Homer or Plato the same path can be followed toward a universe of secular culture whose works, however useless to the salvation of believers, can nevertheless contribute to their moral edification or to their honest amusement. Thus criticism's field of action broadens: one now strives to restore the ancient world in its entirety (by means of the study and collation of texts, inscriptions, buildings, statues, medals), partly because in matters of taste, one grants it an authority analogous to that of the Scriptures in matters of faith; partly also because restoration has an appeal of its own—just as cosmographers assign all the places on earth a real locus in space, so it is similarly fitting to assign events, famous men, inventions, etc. their real date in time, their right place in the course of centuries.

The curiosity of the Renaissance arouses a great influx of information—an influx which entails a risk of disorder against which the mind defends itself with all the resources that might positively contribute to the maintenance of order: chronology, encyclopedic classifications, etc. A sustained exercise of criticism as applied to facts will be necessary before the time when, out of the compilations and the monographs that merely display scholarly competence, the historical work as such will be born, seeking in events the law of their sequence and the possibility of a greater understanding of the present moment. Criticism (particularly as practiced by Pierre Bayle) discovers history not in the narration, but in the very act of situating such historical material as has been verified by the best available resources. The collation of evidence, the rejection of all opinions adopted solely on the guarantee of a so-called trustworthy source, the battle against false conclusions, paralogisms whose theses are supported only by dogmatic tenets: such are some of the aspects of that extirpation of errors which constitutes the official intention of Bayle's *Dictionary*. The field of investigation consists of historical facts and of the contradictory language that relates and organizes them: none of these facts can be declared insignificant, that is to say, none of them is unworthy of restoration to its right place, right time, and right aspect (inasmuch as it is possible to arrive at certainty). Whereas mathematical truths, derived from axioms posited by the mind, do not allow man to leave his mental universe or afford him any opening onto the exterior world, historical certainty aims at a reality wholly exterior to the observer and linked to human action:

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

> I maintain that historical truths can be pushed toward a level of certainty less questionable than that which can be achieved by geometrical truths; provided of course that both kinds of truth be considered according to the kind of certitude that is appropriate to them. . . . Besides, and with all due respect to our worthy mathematicians, it is more difficult for them to achieve the certainty they need, than it is for historians to achieve the certainty that suffices them. No valid argument will ever be brought against the factual truth that Caeser defeated Pompey. And whatever the principles used in the argument, it will be impossible to find anything more undisputable than this proposition—*Caesar and Pompey did exist and therefore they were not just a mere modification of the minds of those who wrote their lives.* But as for mathematics, it is not only very difficult to prove its existence outside our minds, but it is also very easy to prove that it can only be an idea of our mind.[6]

Here Bayle opposes the "factual truth" obtained by historical criticism to the construction through which mathematical discourse is developed. But if he cares about the establishment of fact, it is not because the latter can be employed within the perspective of a broad historical narration based on the ensemble of verified facts, but rather because the critical act has a value of its own. As an act of reason and because of the evidence thus attained, it must elicit the assent of all those who do not wish to remain imprisoned by prejudices; it implies an exigency of universality; and as an act directed against the presumption of knowledge, it renders fragile all dogmatic interpretations of universal history that make Providence reign over events that are often fabricated out of some need of cause: to demonstrate the inanity of alleged facts is also to weaken the credibility of the Cause that has been imposed on them. The mass of duly verified facts does not allow for such an easy ordering: history is perhaps the reign of Evil and this is why, after having clarified the facts, it is better to leave them to their apparently senseless juxtaposition. Besides, the mistakes of others revealed by our own critical eye, make us only all too aware of our own capacity for error. To pursue "factual truths" is to discover how disposed the human mind is to disguise, prevarication, and self-delusion. The ultimate aim of criticism is to change the tone of our quarrels, in which we often speak too loudly only because we do not know what we are talking about. The smallest factual errors deserve to be singled out by criticism:

> The greater the number of falsehoods we can gather on each subject, the better we shall be able to teach man how to know his weakness, and the better we can show him the prodigious variety of mistakes he is susceptible to. We shall be able to make him feel all the more clearly that he is the mere toy of malice and ignorance; that the one picks him up when the other has cast him aside; that if he is clever enough to recognize a lie, he is evil enough to repeat it against his conscience, or that, if he is not evil enough to perpetrate a lie, he contains too much darkness to see the truth. . . .[7] Many people can morally profit from the reading of a large collection of well-established historical falsehoods, if only to become more circumspect in judging their neighbor, and more capable of avoiding the snares that satire and flattery everywhere set against the poor reader. And is it not worthy to correct the nasty inclination we have for rash judgments? And is it not worthy to learn not to put so much stock on the printed word? And is it not a sign of prudence to be discriminating in one's beliefs?[8]

The critical act directed against "falsehood" provides, through the occasion offered by books and libels, a salutary shock which enables one to get rid of a nefarious dependence, though with no guarantee against future relapses into

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

subservience. The distinction between *fact* and *right* [*droit*]—and this should be emphasized—here plays a capital role for, in the domain of right, as Bayle clearly suggests, "truth" (short of lengthy discussions) is ultimately undeterminable, and with the help of philosophical and religious arguments, anyone can plead at length for the superiority of his own convictions. On the other hand, in the realm of *fact* one can always call upon the principle of contradiction whereby "true" and "false" cannot be indifferently affirmed, and whereby any mind receptive to evidence must agree with a fair demonstration. Here, the critical act presupposes the art of isolating the realm of fact. Added to this initial process of segregation will be the disjunction of true from false through an appropriate technique of collation and verification of sources, a technique which, according to Bayle, is invested with rational universality:

> When our only aim is that of gathering factual errors, it is reasonable to assume that each reader follow the same principles and that nobody will accept that which is presented to him as false; for the proofs of factual falsity are not the prejudices of a nation or of a particular religion, but rather maxims common to all men.[9]

But when authors or religious traditions are imprudent enough to found a whole edifice of incontestable truths [*vérités de droit*] on questionable facts, the demonstration of factual falsity, undermining the very basis of the edifice, will have considerable consequences: for criticism this destruction will be a more distant result which will only complete the daring pedagogy of intellectual prudence. We witness the disappearance of a major object of belief and the advent of a new one, infinitely more modest (but also more "real" and more worthy of confidence)—and at the same time the rational mind becomes aware of its power to effect this substitution, to provoke this change, this "revolution." Thus the criticism of factual errors becomes instrumental in a *general criticism*. In a wider sense, the restitution of the precise fact to the limited authority of its truth entails the destitution of the philosophical or theological discourse whose claim to unlimited authority was based on usurped premises. Even though, according to Bayle, human reason (which has produced and tolerated so many errors) is feeble, the critical act is the work of that reason—but of a reason now conscious of its fallibility and henceforth aware of what it can affirm and refute without error.

The fact is therefore simply the occasion of that act which restores it to its "factual truth." Because Bayle is not trying to construct a history, facts are not for him the raw material for some ulterior work. The problem of the choice of significant facts, of the selection of the "data" according to the values and interests of the historian, is known to everyone today, for the facts that provide the basis for an interpretation seem to result themselves from a previous interpretation. But Bayle restricts himself to the facts themselves without trying to organize them into a narrative or explanatory discourse which would interrelate them. His confessed aim is simply that of founding, within the context of the knowledge of things human, a domain of evidence accessible to all men of good faith: thus in the midst of controversy and despite the conflict of confessions and the confused jumble of heresies, a meeting ground is established, a truce zone. In this zone, small but solid, all contenders can agree and conclude an armistice which, with their consent, will turn into a general peace: they will not be asked

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

to renounce their convictions, but to reserve them (on condition of reciprocity) in their heart of hearts.

Thus, paradoxically, the establishment of a realm of factual objectivity guarantees a status of immunity to subjective conviction, whereby its own right is affirmed once and for all, sheltered from any dogmatic authority that might pretend to rule over its "conscience." Individual faith is protected from all threatening intrusion because fact alone can be verified and thus become, in its own essence, the occasion for general agreement—that is, if men are willing only to realize the advantages they can collectively reap from this. The advantages themselves can be demonstrated as a fact, and politicians are well advised to take this into consideration. In several respects this clear-cut distinction between factual truths and objects of faith already establishes the dividing line which will separate the domain subject to the authority of the State (neutral and secular) and the realm of private conviction, inaccessible to all intrusion on the part of an external power.

Thus, by firmly establishing the reign of an objectivity recognizable by all, Baylian criticism manages to accommodate a subjectivity that wishes to depend exclusively on itself. The control exercised by criticism on those historical foundations claimed by the spiritual power (itself traditionally considered as the basis of temporal power) ends up disarming those interventions which could have threatened all freedom of thought in the name of religious orthodoxy. In this particular case, the trick of criticism consists in disguising itself as an act of restitution (only texts that are not suspects and proofs that are solid will be allowed to survive); but such criticism, constantly demanding positive proofs and reliable texts, will leave nothing standing and will hence function as an act of destitution. The authority of the individual conscience is thus reinforced and comforted until, toward the end of the eighteenth century, it becomes the main recourse for theoreticians of political power who, unable to appeal to the authority of revelation or of tradition in order to establish public order, will have to invoke the testimony of free inner conviction as sole guarantor of respect for law and of civic duty. Liberal individualism, with all the leeway it grants to the personal interpretation of inner authority, finds its safeguard in the criticial activity that submits to strictest examination those arguments of authority and antiquity on which absolutist dogmatism had depended.

The full scope accorded to the self, to subjective experience, will not necessarily entail a feeling of gratitude and solidarity toward a historical criticism concerned with the texts of the past and with the careful, objective reassessment of verified facts. No sooner is subjective activity liberated that it tends to forget how much it owes its immunity, its promotion to the rank of supreme authority, to those critical labors which have terminated the reign of extrinsic authorities; instead of acknowledging its debt to the activity of restorative criticism, the autonomous conscience seems aware only of what mutually differentiates them. It is obvious that this exuberant conviction, given its poetical and often exalted character, seems hardly compatible with the vigilance of grammarians often bogged down in thankless and apparently futile philological tasks.

Since the beginning of the eighteenth century, it has been quite fashionable to quarrel with erudition in the name of reason or of the creative imagination, as if it were a question of vanquishing some rival. To be sure, the scholar does not seem to have anything in common with the inspired writer or with the specula-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

tive philosopher or experimental scientist: their methods and aims could not be more at variance. But what has frequently remained misunderstood is that the work of historical criticism (grammatical and philological) not only presupposes the same legitimacy granted to the autonomous exercise of reason, but also does its best to prevent all spiritual and temporal powers (the Church and the police) from asserting themselves over what poets, philosophers, and physicists consider as their own truth. Historical and philological criticism have therefore not only been the school of free judgement, but also the victorious avant-garde that first silenced the power of dogmatism, thus paving the way for those individuals determined to listen only to their own convictions and to their own personal reason.[10]

II

This much said, it might appear as if scholarly criticism had only a historical role to play at a preliminary, preparatory stage: to clear the way, to restore ancient texts to their authentic state, to assure the transmission of power between extrinsic authority (the Church, the Revelation) and inner authority (conscience, private conviction). This, at any rate, is the way d'Alembert, the *philosophes*, and the principal figures of the *Encycopédie*, saw it. But we know that philology and historical criticism made great strides in the nineteenth century. And we know that it was only then that many philological disciplines attained their full scientific status. This obviously entailed far more than a mere survival of that scholarly curiosity which the theoreticians of 1760 already considered as obsolete (because its fruits had already been harvested).

At the risk of simplifying, it should be recalled that during the nineteenth century scholarly criticism was called upon by two systems of thought than had emerged within Western culture. The first one, related to the "neoclassical" movement, wants to reestablish vis-à-vis antiquity, and particularly Greece, a relationship at once more intense and more exclusive than that of the previous centuries: the individual, freed from the dominion of religious dogmas and having (in Kant's phrase) at least reached his majority, seeks the model for his authentic being in the works of that emergent civilization. The quest for the true face of the Greek world—which must at once be definitely exhumed and freed from the faulty interpretations accumulated over the course of centuries—demands the deployment of all the resources of criticism. For philology, for hermeneutics, the aim of the quest is very clearly defined: to recover, beyond any risk of illusion or error, some of the original insights of the human conscience.

It is quite clear that as a technique, philology is considered as that tool which alone can provide access to a subjective experience that may be obsolete, but is no less precious for that. (For sure, classical philology, through the development of its own tools—above all those of linguistics—ran the risk of forgetting this finality, of merely cultivating scholarship for its own sake, of floundering in the accumulated mass of partial and limited details.) But a new displacement of authority was not long in happening. In the first place, as was inevitable, the men of the nineteenth century were doomed to realize that they were incapable of reliving the Greek ideal. And to some this earliest spiritual universe to

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

emerge in Europe no longer seemed the site of some unsurmountable perfection. Within the perspective of a philosophy of history convinced of progressive perfection in time, the civilizations of the past merely bear testimony to a yet incomplete awakening, to a rough outline of an intuition destined to transform and purify itself. For this philosophy of progress, philology and scholarly criticism have a capital importance analogous to that which evolutionary biology will attribute to paleontology and embryology: the knowledge of the organisms of the remote past, the understanding of formative processes enable one to trace the vectors of evolution and to anticipate those future horizons toward which all progress tends. At the same time, the authority to which all critical judgment lays claim is no longer solely limited to the individual conscience; the collective becoming [*devenir*], initially viewed by criticism simply as an object of study, now tends to assume divine attributes: soon the theoreticians of progress will endow this newly divinized humanity with a function of authority analogous to the one which the older dogmatism had attributed to the Revelation and the Scriptures.

This change happened in stages. And one of these stages is certainly marked in *l'Avenir de la Science*, the book Ernest Renan wrote in his youth, in 1848, and published only on the eve of his death in 1890. In *l'Avenir de la Science*, Renan was able to see very clearly the instrumental role played by philology in the Reformation, that is to say, in the movement aimed both at the restoration of the Scriptures and at the justification of the personal reading that each conscience could draw from them. "The Reformation was born in the midst of philology." So the nineteenth century, which aspires to be a critical century, must similarly elevate philology to the rank of master science.

> Universal criticism is the only sure characteristic that can be assigned to the delicate, fleeting, elusive thought of the nineteenth century. . . . Criticism is therefore the form through which the human mind exercises itself in all its endeavors; now, if criticism and philology are [not] identical, they are at least inseparable. To criticize is to posit oneself as spectator and judge amid the variety of things; now, philology is the interpreter of things, the means of establishing communication with them and of understanding their language. The day that philology dies, barbarism will be reborn, and credulity will again become master of the world.[11]

For Renan, philology, "the exact science of things of the mind," leads to the "direct and pragmatic knowledge of things." Now, the knowledge of things is none other than the knowledge of their history, of their becoming, and of that causal law—the only unchangeable element—which induces change. For instance, the knowledge of languages is the knowledge of their evolution: "History is the true form of the science of languages." The individual can understand himself only in his becoming [*devenir*]: "The soul is individual becoming, just as God is universal becoming. . . . The science of man will not be properly clarified until we are convinced that conscience makes itself, that, initially weak, vague, and uncentered both in the individual and in humanity as a whole, it proceeds through diverse phases until it attains its plenitude. Then we shall understand that the science of the individual soul is the history of the individual soul, and that the science of the human mind is the history of the human mind."[12]

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Now if, as Renan affirms elsewhere, a permanent law expresses itself through this universal becoming, it obviously follows that the same laws govern the formation of languages and their transformations, which is nothing other than the life of collective consciences marching toward their future fulfillment: "The creative spirit of language is present in that which presides over its revolutions: for the force that gives life is, in the end, the force that gives birth, and to develop is, in a sense, to create."[13] By attributing the name of God to this universal becoming, Renan submits God to the law or system of laws which act immutably upon the fluid succession of phenomena. Thus the project of a collective history of humanity, of an "embryogeny of the human mind,"[14] aims, as does its remote conclusion, at the total contemplation of the "system of things." The final vision promised to a humanity enlightened by Science will be the collective knowledge of the collective becoming as manifested in its ultimate truth. The self which elaborates and enjoys knowledge is, therefore, the self-produced collective self: the knowledge it desires is imbued with itself, being the recapitulation of its own history. History completes itself in the knowledge of history, and the structure now grasped achieves the fixed brilliance of a diamond: ". . . I feel that if I had ten lives to live simultaneously, so as to explore every world, myself there at the center, inhaling the scent of all things, judging and comparing, combining and inferring, I would finally get to the system of things. Well, what no individual is capable of, humanity will do: for it is immortal and we all work for it. Humanity will be able to *perceive* the true physiognomy of things, that is, it will attain truth at all levels. Will you then say that those who have contributed to this immense work, who have polished one of the facets of this diamond, who have rid it of some of the dross that dims its native brilliance, are just pedants, idlers, ponderous minds busy wasting their time?"[15]

Although the totality of meaning, that asymptotic object of completed knowledge, might seem the opposite of the minute "fact" isolated by the philologist, it is nevertheless here, in the fine detail which the scholar discerns and adds to other analogous discoveries, that the former kind of knowledge has, according to Renan, its starting point. It is from this fragmentary starting point that inductive work commences its course through the universality of the "sciences of man." The final "diamond" remains, of course, a hope that requires an act of faith, like the promise of Paradise: history will culminate, for everyone, in intelligent plenitude. But what is absolutely sure is that, in order to approach this future, one must begin by recovering the memory of the past—and therefore devote much time to thankless tasks, to the strenuous penances of philology.

What Renanian criticism tries to reestablish is the truth that manifests itself at the level of the collective becoming (through languages, beliefs, institutions); and, in order to attain this, any material will do, provided that it is scrupulously assessed and verified.

In practice, the philological exploration of documentary material could be a never-ending process, never allowing more than a provisional overview. One must attempt to organize incomplete items into intelligible syntheses, being always aware that they will be valid only until further information is supplied, and that they are admissible only "given the present state of knowledge." This is what Renan does when writing *The Origins of Christianity*.[16] He not only knows

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

that his work is wholly dependent on the nature and the current number of his sources, but he is also too intelligent not to be conscious of the emotional interest that has involved him in his work and that risks, at any point, to falsify its results. For the suspicious reader, it would not be difficult to go yet a step further and to denounce the "ideological" character of such an enterprise: it is merely a former seminarian's way of salving his conscience after having turned away from Christianity. Studying the historical genesis of the faith he abandoned, Renan finds ample reasons to contest any idea of revelation: for the careful student of its birth, the Christian faith cannot avail itself of any argument that would advocate submission. If the act of faith wishes to remain alive, it will have to turn toward the spiritual becoming of humanity as a whole. If asceticism is not to disappear, it will be exercised in the labors of scholarship. Always on the lookout, the critical mind is well aware of the conjectural element entailed by such a faith: God is shoved toward the future, which allows acceptance of a present less marked by the imminence of the divinity than by its absence. Faith is not only displaced toward another object, but because this object is out of reach, the act of faith is itself postponed and turned into nothing more than a probable hypothesis. Philological meticulousness derives its dignity and its justification from providing raw material for some glorious future knowledge, but within the imperfect state of the present, this conviction is only a convenient and makeshift illusion.

For Renan, the perpetual alertness of philological attention maintains its old antidogmatic value: for, if it gathers the facts on which the grand totalizing induction will ultimately be built, philology retains that polemical function it had for Pierre Bayle; it is still a weapon drawn against any usurpatory authority that demands the abdication of reason. In theory, Renan sees the individual disappear within the collectivity, but he reserves for himself an exigency that closely ties criticism to individual authority: criticism is the observing individual who will never consent to be stripped of his right to free examination. The total knowledge that Renan ascribes to the humanity of the future is the imaginative expansion, the generalized plenitude of a pleasure that refers back, above all, to "myself there at the center."

Thus, between a vigilant philology that clings to minute details, and an intelligible totality of whose remote and hypothetical character Renan is perfectly aware, there is room for a wisdom that is at once broad, sceptical, tolerant, a wisdom that knows that faith is no longer (or not yet) possible, and that science has not yet come to its logical conclusion.

With Renan, this appeal to the future never assumes the character of a dogma. The diamond of ultimate knowledge shines at the fleeting point of all partial histories, of all segmentary evolutions, but Renan hopes without promising anything, and above all forbids himself to anticipate anything. The exact knowledge of the past is for him the condition sine qua non of all future accomplishment. Philological knowledge is therefore not only a necessary instrument, but also a portion of that knowledge to come; to give it up, in spite or in impatience, would mean to lose any chance of attaining it.

But this appeal to the future may assume another aspect, as recent tendencies and contemporary attitudes clearly testify. Our discussion of this point will

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

be limited to a rather schematic overview. The advent of the future collective good, initially derived from a progressive theory of history, frees itself from its empirical premises to become a nonappealable postulate. Whatever name we give it, this future point has often been compared to the glorious future promised by religious eschatologies and doctrines of salvation. But here an essential difference must be pointed out: religious eschatologies, although opening the space of the future time, based their legitimacy on an ancient word: revelation, scripture, covenant, or promise. To sum up in a few words, the belief in a heavenly Jerusalem implied a belief in the story of Genesis, of the Fall, in the genealogy of David, and more generally, in the supernatural provenance of the Book in which these events of the past were recorded. As consummated facts or texts, both sacred history and the Scriptures were liable to critical verification: if the alleged chronology was no longer admissible, or if it appeared that the "revealed" text was not of a single piece, the authority of the dogma was shaken. And as soon as the past was no longer certain, the future was no longer assured.

If the account of the beginning of time was no longer believable, it becomes useless for man to subordinate his existence to an end of time defined by a strict correlation with the discredited account. The old dogmatism was therefore vulnerable to criticism because of the image of the beginning that it had tried to impose. (This is how Voltaire fought Christianity, by attacking the Old Testament). The new dogmatisms, whether they appeal to individuals or collective faiths, seem to have tried to escape this kind of strict verification. The authority that they seek to impose is entirely gathered in the future: it is situated beyond history and sees itself as the "end of history." As a result, it is invulnerable to any historical, philological, or scholarly refutation of an event located at the beginning of history or even in contemporary history.

Modern dogmatic eschatologies take historical events only as examples or pretexts, never as foundations for the legitimacy of belief. Belief, oriented entirely toward a time that is yet to come, offers no target to "factual" criticism—which is not to say that the adepts of the new authority have no need of a suitably reconstructed image of the past, but the scientific challenging of this image does not imply for them the collapse of faith. In their eyes, the image of the past can be manipulated without damage as long as the authority attributed to future beatitude is not diminished; on the other hand, they are the first to revise the account of past events as often as necessary when it is a question of protecting the determining importance of that future to which they claim to subordinate everything. The new dogmatisms sway between tolerance as concerns facts related to the past of the species, and extreme intolerance as concerns the constantly remanipulated contemporary hagiography needed by mobilizing propaganda. In the battle for what is believed to be collective salvation, as in the appeal to religious conversion, all that does not contribute to the *unum necessarium* is declared frivolous, irrelevant, guilty. Therefore only a militant criticism submitted to eschatological authority is judged acceptable. And because the older dogmatisms were likely to be attacked in the proofs they alleged, criticism was able to forge its weapons into a technique for the verification of textual proofs. But the new dogmatisms generally require that we believe their promises with no proofs other than the dissatisfaction (which they try to aggravate) that we feel with the present state of the world. They invite us to draw a blank check—an uncovered check—on the future. For, according to their system of

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

interpretation, itself supported by the image of a glorious and unverifiable accomplishment, even their failures and their errors are indicative of the overall infallibility of the doctrine.

On the other hand, lacking such assurance, philological criticism and scholarly history (and by this I mean history based on the widest accessible documentation) cannot avoid some pangs of conscience when they sense themselves under attack. These disciplines possess many scrupulous scholars, carefully accomplishing their specialized tasks to the best of their capacities, who are left speechless when people who know beyond the shade of a doubt what we can expect from the future ask them *what is the purpose* of their work. They dare not answer, quite simply, that they are doing what they please and that only violence could forbid them to continue. Generally they allege the purity and disinterestedness of their research, which is not the best justification. Declared either obsolete or sterile, or qualified as individualist or bourgeois, their disciplines will not be able to defend their validity by declaring themselves impartial and neutral, as if unwilling to prejudice any choice in advance. In fact, it will not be easy for historical relativism to have them admit that they depend on a historical moment, that they derive from a bias, and that they are tied in to interests.

Scholarly disciplines have everything to gain by becoming aware of their implicit biases and by confessing their (self-) interest—about which they have absolutely no reason to blush. They have, both in their past as well as in their present, solid proofs they can invoke to show that they have been suspected only when all aspects of the free exercise of reason became suspect; that their impartiality and their apparent disinterestedness, far from hesitating before historical choices, are the direct consequence of an option whose range goes far beyond the simple principles of methodological rigor implemented within an isolated discipline; that if they have all too often closed themselves off within the narrow domain of positive inventory and verification, it has been only because they had forgotten that, aiming at exactitude of detail, their finality is first of all the pursuit of sense, and that in this stubborn, painstaking quest, they are also, like eschatological beliefs, oriented toward the future. But they are oriented toward a different future; they aim at another tomorrow, that which manifests itself as the reward of correct interpretation and which is therefore constantly elusive, for there can never be any exhaustive interpretation. Such a future claims no constraining authority a priori, nor does it authorize any imposition of constraint. Reason, qua will to examine, begins by denying all criteriological authority to those extrapolated figures of desire which are projected toward the culmination of history. It simply invites a common quest according to the rules of critical ethics: the only future that it can conditionally promise—for it is not sure it will come true—is the one in which more and more individuals meet in respect for demonstrated truths, in pursuit of that which has not yet been approached, and in the refusal of usurped authority. This particular future, among all those which, given the present state of the world, may happen, might not seem very exciting to naive minds, nor does it exercise a very great power to mobilize men. I hope, nevertheless, that this will not cease to be a *possible* future. For if this possibility were to vanish, it would be the sure sign of some lasting calamity.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Let us recapitulate. Initially, the ambition of critical scholarship was that of restoring the full force of authority to the revealed Word or to the great literary models of the past. But the difficulty of the task, the obstacles, the irreducible uncertainty of the sources, the conflicts among commentators, as well as the growing doubt as to the legitimacy of the authority thus served, forced critical reasoning to retreat. This retreat was turned into a victory when the critical mind discovered in itself the authority, the sovereign reason, which it had refused to consider as imposed from the outside by the infallible text of some past revelation. And this triumph of the *analytical mind* (*esprit d'examen*) went hand in hand with the appeasement of religious conflicts and with the end of antagonism between rival theological dogmatisms. Reinforced by this victory, critical knowledge was soon to elaborate its own myth and to turn this into a substitute dogma: historical scholarship, as Renan will affirm, is one of the aspects of the progress of reason, at the end of which humanity, conciliated and divinized, will enter into the great unitary realm of achieved Science. We have shown how this new dogmatism, by transferring the emergence of authority to some remote future, can become (despite the best intentions of its adepts) as tyrannical as those orthodoxies derived foom sacred ancient scriptures, because it is immune to inquiry, to doubt, and to the very test of criticism: its interpretation of past events can be judged only in terms of a hope that tries to hasten its own realization by assuming the role of law. No one will hence be surprised if this projection into the future of an act of faith that is fundamentally religious in nature provokes, in its turn, a reaction of agnosticism (as was the case with Renan). Critical reason discovers itself sinning against its own rules whenever, dreaming of its own apotheosis, it turns this into the theme of a new eschatology. And just as the ancient Revelation saw itself as a Kingdom to come, the polemic against revealed dogma transforms itself into a polemic against those utopian systems which try to read the meaning of the past in the mirror of some radiant future.

Of course, as in the age of Pierre Bayle, this critical endeavor could have an ironic impact insofar as it could confront utopias with what men have actually undertaken in their name: it would show to what extent the new social orders differ from the image originally envisaged. (And yet Utopia can defend itself far better than Revelation: it simply takes shelter in some new future and claims to be still at the preliminary stages of its realization.)

It is true that criticism does not adapt itself very easily to this ironic function in which it can play only a subordinate role: a role of destitution vis-à-vis usurped authorities; a role of conciliation among systems that claim authority; a role of surveillance, mediation, and assistance. Why could not criticism, in and of itself, become a sovereign language [*parole*], a definitive verdict, an ultimate trial?

Today authority is a privilege that pertains almost exclusively to scientific discourse. Renan had sensed this back in 1848, but he imparted to science the function of an inverse theology in which God was expected to manifest himself at that moment when scientists had finally completed their task. In our century, the "human sciences"—sociology, psychology, linguistics—have almost come to share the prestigious status of the natural sciences, which should explain why one of the final metamorphoses of criticism should involve exchanging the old philological approach for the more authoritative style of the "human sciences."

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

For philology, deprived of any hope of ever acceding to the absolute, finds little satisfaction in its own praxis: nothing seems to confer on it the general validity, the power of expansion, that characterize the dominant discourses. Reduced to its barest self, philology appears as a mere discipline of information, respectable to be sure, but devoid of finality. By reabsorbing itself into the "human sciences," by almost melting into them, it seemingly renounces its precise, limited problems: it leaves the critic free to establish laws, relations, structures, and systematic formulas—whose profits can be entered under the column of *general theories* of culture, literature, etc.

In this respect recent history is quite illuminating. From structural analyses we have passed quite rapidly to global theories of literature; all debate has been shifted to these theories, and the discussion is interminable. And we know by now that this eternal discussion has also lost all pertinence; it no longer grasps anything. Much unlike what happens in the natural sciences (except in the speculative regions of physics), these discussions, lacking any boundary imposed by a precise object, end up bearing on the formal configurations of the theories themselves, which are, in their turn, indefinitely modifiable and incapable of discerning any ultimate proof of their validity. And the authority to which they pretend again dissolves like smoke.

Formulated after a far too cursory contact with the text, theory must realize that the farther it gets from it the weaker it becomes: a return to the text is necessary if theory is to survive. And when this return takes place, critical theory will have to rediscover the justification of philology—insofar as philology, which is first of all the science of the establishment of texts, is always the science that must answer the fundamental question of the nature and the meaning of interpretation.

The concept of interpretation to which we have now returned almost automatically forces us to reconsider the notion of authority as it applies to the knowledge of cultural facts. And my attention focuses on those points where, to our demanding minds, authority simply vanishes. I am mainly concerned with two of them, both opposed to each other, and almost symmetrical: the first is located at the level of the establishment of particular facts, when the exercise of scholarly research limits itself to the art of verification and statement: the number of those concerned with this kind of demonstration is generally so small that it is not surprising that, outside a small circle of specialists, one finds only respectful indifference. Conversely, when rival theoretical discourses with universal pretentions proliferate, the very indefinite character of their conflict fragments authority, limits it, weakens it, and finally reduces it once again to a small group of adepts and to the limited duration of a vogue. In the first case, we have a series of "partial" objects that are indefinitely multipliable; in the second, we confront a series of interpretive systems which are also indefinitely multipliable insofar as their "practical" applications have not been successfully tested and countertested.

Shall I add that I reject neither the work that goes into the scrupulous establishment of "facts" nor theoretical speculation? I merely wish to point out the relative sterility of pursuing these two goals separately. But everything can change if "fact" meets with an interpretive reading, and theory returns more closely to fact and consents to adapt and modify itself according to the demands

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

of its object, so as to increase its textual understanding and comprehensiveness. The central locus of authority—such as I, along with many others, am inclined to recognize it—rests at the meeting point of "facts" (dependent on our subjective choice, confirmed and verified by a series of objective procedures) and of our "theoretical" inquiry. The interpretive act achieves complete validity only when an energetic question, with universal implications, bears directly on an object that is clearly grasped in its singular profile. From such a meeting, which is a task with no a priori guarantee of success, some meaning may be gained. And the meaning thus achieved is the only thing that can speak to us with authority; but this evidence does not await us as some preexisting secret or some veiled message: it is the result of an act, the stake of a bet. It is therefore not a question of perceiving an absolute statement hidden behind appearances, but rather of acknowledging those appearances, of showing them our greatest consideration, so that we might, with their consent, build in (and out of) them a web of significant relations which, far from being answers, will be able to transform our earlier questions into new queries. The interpretive act involves, at once, deciphering and creation, philological patience and productive invention. If today the interpretive act is no longer subordinated to some absolute message or to a Scripture which it must restore and explicate; if, by establishing relations, it works within the realm of the relative, it is no less true that its very relativity is the only bearer of authority in the eyes of a justifiably sceptical reason that insists on verifying all claims to power.

Translated by A. Cancogni and R. Sieburth

References

[1]This study should be pursued in the various Western languages where the term has been in use. One will notice that our investigation is restricted to the French sphere. I have outlined several more general fundamentals in "On the Fundamental Gestures of Criticism," *New Literary History*, 5 (Spring 1974): 491–514.

[2]M. Starobinski's word is *restitution*—which we have chosen on the whole to translate as "restoration," although in the context of philology, the "establishment" or "reconstitution" of texts would be equally apt.—*translators*

[3]From the article "Critique" in the *Encyclopédie ou Dictionnaire raisonné des Sciences, des Arts et des Métiers*, vol. 17. This text can also be found in Marmontel's *Elémens de Littérature* (6 vols., Paris, 1787), vol. 2, pp. 221 ff.

[4]Ibid.

[5]This argument appears in all its clarity in Book VI of Rousseau's *Emile*, the famous "Profession du foi du Vicaire Savoyard."

[6]"Projet d'un dictionnaire critique," in Pierre Bayle, *Dictionnaire historique et critique* (4 vols., Rotterdam, 1720), vol. 4, pp. 2983–2984.

[7]Ibid., p. 2984.

[8]Ibid., p. 2983.

[9]Ibid., p. 2984.

[10]For all these points consult Reinhart Koselleck, *Kritik und Krise* (Freiburg-München: Suhrkamp, 1959, 2nd ed. 1973).

[11]Ernest Renan, *L'Avenir de la Science* (Paris, 1890), pp. 145–146.

[12]Ibid., p. 182.

[13]Ibid., p. 171.

[14]Ibid., p. 164.

[15]Ibid., p. 148.

[16]Ernest Renan, *Histoire des origines du Christianisme* (8 vols., Paris 1863–1883).

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

STEVEN WEINBERG

The Search for Unity: Notes for a History of Quantum Field Theory

Introduction

QUANTUM FIELD THEORY is the theory of matter and its interactions, which grew out of the fusion of quantum mechanics and special relativity in the late 1920s. Its reputation among physicists suffered frequent fluctuations in the following years, at times dropping so low that quantum field theory came close to be abandoned altogether. But now, partly as a result of a series of striking successes over the last decade, quantum field theory has become the most widely accepted conceptual and mathematical framework for attacks on the fundamental problems of physics. If something like a set of ultimate laws of nature were to be discovered in the next few years (an eventuality by no means expected), these laws would probably have to be expressed in the language of quantum field theory.

To the best of my knowledge, there does not exist anything like a full history of the past fifty years of quantum field theory. Existing histories of modern physics cover special and general relativity pretty thoroughly, and they take us through the early years of quantum mechanics, but their treatment of quantum mechanics generally ends with the triumph of the statistical interpretation around 1927. This is a pity, not only because of the fundamental nature of quantum field theory, but also because its history offers an interesting insight to the nature of scientific advance.

It is widely supposed that progress in science occurs in large or small revolutions. In this view, the successes of previous revolutions tend to fasten upon the scientist's mind a language, a mind-set, a body of doctrine, from which he must break free in order to advance further. There is great debate about the degree to which these revolutions are brought about by the individual scientific genius, able to transcend the fixed ideas of his times, or by the accumulation of discrepancies between existing theory and experiment. However, there seems to be general agreement that the essential element of scientific progress is a decision to break with the past.

I would not quarrel with this view, as applied to many of the major advances in the history of science. It certainly seems to apply to the great revolutions in physics in this century: the development of special relativity and of quantum mechanics. However, the development of quantum field theory since 1930 provides a curious counterexample, in which the essential element of progress has

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

been the realization, again and again, that a revolution is unnecessary. If quantum mechanics and relativity were revolutions in the sense of the French Revolution of 1789 or the Russian Revolution of 1917, then quantum field theory is more of the order of the Glorious Revolution of 1688: things changed only just enough so that they could stay the same.

I will try to tell this story here without using mathematics. To some extent, I will let the history do the job of explication—I will go into some of the historical developments more fully than others, because they help to introduce ideas which are needed later. Even so, this is not an easy task—there is no branch of natural science which is so abstract, so far removed from everyday notions of how nature behaves, as quantum field theory.

This cannot be a story of physics in one country. As we shall see, quantum field theory had its birth in Europe, especially in Germany and Britain, and was revived after World War II by a new generation of theoretical physicists in Japan and the United States. The United States has been somewhat the center of the intense activity of the last decade, but physicists from many countries in Europe and Asia have made essential contributions. And although there are discernible national styles in physics, they have played only a minor role in this history. It is not the national or the social or the cultural setting that has determined the direction of research in physics, but rather the logic of the subject itself, the need to understand nature as it really is.

This article is not a history of quantum field theory, but only "notes for a history." A great deal of work needs to be done by professional historians of science in uncovering the story of the last half-century of theoretical physics. I would be delighted if this article were to spur someone to take on this overdue task.

Prehistory: Field Theory and Quantum Theory

Before quantum field theory came field theory and quantum theory. The history of these earlier disciplines has been told again and again by able historians of science, so I will do no more here than remind the reader of the essential points.

The first successful classical field theory was based on Newton's theory of gravitation. Newton himself did not speak of fields—for him, gravitation was a force which acts between every pair of material particles in the universe, "according to the quantity of solid matter which they contain and propagates on all sides to immense distances, decreasing always as the inverse square of the distances."[1] It was the mathematical physicists of the eighteenth century who found it convenient to replace this mutual action at a distance with a gravitational field, a numerical quantity (strictly speaking, a vector) which is defined at every point in space, which determines the gravitational force acting on any particle at that point, and which receives contributions from all the material particles at every other point. However, this was a mere mathematical *façon de parler*—it really made no difference in Newton's theory of gravitation whether one said that the earth attracts the moon, or that the earth contributes to the general gravitational field, and that it is this field at the location of the moon which acts on the moon and holds it in orbit.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Fields really began to take on an existence of their own with the development in the nineteenth century of the theory of electromagnetism. Indeed, the word "field" was introduced into physics by Michael Faraday in 1849. It was still possible for Coulomb and Ampère to consider electromagnetic forces as acting directly between pairs of electric charges or electric currents, but it became very much more natural to introduce an electric field and a magnetic field as conditions of space, produced by all the charges and currents in the universe, and acting in turn on every charge and current. This interpretation became almost unavoidable after James Clerk Maxwell demonstrated that electromagentic waves travel at a finite speed, the speed of light. The force which acts on electrons in the retina of my eye at this moment is not produced by the electric currents in atoms on the sun at the same moment, but by an electromagnetic wave, a light wave, which was produced by these currents about eight minutes ago, and which has only now reached my eye. (As we shall see, an attempt was made in 1945 by Richard Feynman and John Wheeler to account for this retardation of electromagnetic forces in an action-at-a-distance framework, but the idea did not catch on, and they both went on to more promising work).

Maxwell himself did not yet adopt the modern idea of a field as an independent inhabitant of our universe, with as much reality as the particles on which it acts. Instead (at least at first) he pictured electric and magnetic fields as disturbances in an underlying medium—the aether—like tension in a rubber membrane.[2] This would have had one obvious experimental implication—the observed speed of electromagnetic waves would depend on the speed of the observer with respect to the aether, just as the observed speed of elastic waves in a rubber membrane depends on the speed of the observer relative to the membrane. Maxwell himself thought that his own field equations were in fact valid in only one special frame, at rest with respect to the aether.[3] The notion of an aether that underlies the phenomena of electromagnetism persisted well into the twentieth century, despite the repeated failure of experimentalists to discover any effects of the motion through the aether of the earth as it revolves about the sun.

But even with the problem of the aether unresolved, the idea of a field as an entity in its own right grew stronger in physicists' minds. Indeed, it became popular to suppose that matter itself is ultimately a manifestation of electric and magnetic fields, a theme explored in the theories of the electron developed in 1900–1905 by Joseph John Thomson, Wilhelm Wien,[4] M. Abraham,[5] Joseph Larmor, Hendrik Antoon Lorentz,[6] and Jules Henri Poincaré.[7] Finally, in 1905 the essential element needed to free electromagnetic theory from the need for an aether was supplied by the special theory of relativity of Albert Einstein.[8] New rules were given for the way that the observed space and time coordinates of an event change with changes in the velocity of the observer. These new rules were specifically designed so that the observed speed of a light wave would be just that speed calculated in Maxwell's theory, whatever the velocity of the observer. Einstein's theory removed any hope of detecting the effects of motion through the aether, and although the aether lingered on in theorists' minds for a while, it eventually died away, leaving electromagnetic fields as things in themselves—the tension in the membrane, but without the membrane.

As it happens, it was the study of electromagnetic phenomena that gave

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

rise to the quantum theory as well as to special relativity. By the end of the nineteenth century, it was clear that the classical theories of electromagnetism and statistical mechanics were incapable of describing the energy of electromagnetic radiation at various wavelengths emitted by a heated opaque body. The trouble was that classical ideas predicted too much energy at very high frequencies, so much energy in fact that the total energy per second emitted at all wavelengths would turn out to be infinite! In a paper read to the German Physical Society on December 15, 1900, a resolution of the problem was proposed by Max Karl Ernst Ludwig Planck.[9] It will be worthwhile for us to concentrate on Planck's proposal for a moment, not only because it led to modern quantum mechanics, but also because an understanding of this idea is needed in order to understand what quantum field theory is about.

Planck supposed that the electrons in a heated body are capable of oscillating back and forth at all possible frequencies, like a violin with a huge number of strings of all possible lengths. Emission or absorption of radiation at a given frequency occurs when the electron oscillations at that frequency give up energy to or receive energy from the electromagnetic field. The amount of energy being radiated per second by an opaque body at any frequency therefore depends on the average amount of energy in electron oscillations at that particular frequency.

It was in calculating this average energy that Planck made his revolutionary suggestion. He proposed that the energy of any mode of oscillation is *quantized*—that is, that it is not possible to set an oscillation going with any desired energy, as in classical mechanics, but only with certain distinct allowed values of the energy. More specifically, Planck assumed that the difference between any two successive allowed values of the energy is always the same for a given mode of oscillation, and is equal to the frequency of the mode times a new constant of nature which has come to be called *Planck's constant*. It follows that the allowed states of the modes of oscillation of very high frequency are widely separated in energy, so that it takes a great deal of energy to excite such a mode at all. But the rules of statistical mechanics tell us that the probability of finding a great deal of energy in any one mode of oscillation falls off rapidly with increasing energy; hence the average energy in oscillations of very high frequency must fall off rapidly with the frequency, and the energy radiated by a heated body must also fall off rapidly with the frequency of the radiation, thus avoiding the catastrophe of an infinite total rate of radiation.

Planck was not ready to apply the idea of energy quantization to radiation itself. (George Gamow[10] has described Planck's view as follows: "Radiation is like butter, which can be bought or returned to the grocery store only in quarter-pound packages, although the butter as such can exist in any desired amount.") It was Einstein[11] who in 1905 proposed that radiation comes in bundles of energy, later called *photons*, each with an energy proportional to the frequency.

In 1913 the ideas of Planck and Einstein were brought together by Niels Bohr in his theory of atomic spectra.[12] Like the hypothetical modes of oscillation in Planck's work, atoms in Bohr's theory are supposed to exist in distinct states with certain definite energies, but not generally equally spaced. When an excited atom drops to a state of lower energy, it emits a photon with a definite

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

energy, equal to the difference of the energies of the initial and final atomic states. Each definite photon energy corresponds to a definite frequency, and it is these frequencies that we see vividly displayed when we look at the bright lines crossing the spectrum of a fluorescent lamp or a star.

This early quantum theory, from Planck to Bohr and for a decade after, was inspired guesswork, ad hoc mathematical manipulation, justified by its brilliant success in explaining the behavior of atoms and radiation. Quantum theory became the coherent scientific discipline known as *quantum mechanics*, through the work of Louis de Broglie, Werner Heisenberg, Max Born, Pascual Jordan, Wolfgang Pauli, Paul Adrian Maurice Dirac, and Erwin Schroedinger in 1925–1926.[13] Armed with this formalism, theorists were able to go back to the problem of determining the allowed energy levels of material systems, and to reproduce the successful results first found by Bohr. But despite its origins in the theory of thermal radiation, quantum mechanics still dealt in a coherent way only with material particles—electrons in atoms—and not with radiation itself.

The Birth of Quantum Field Theory

The first application of the new quantum mechanics of 1925–1926 to fields rather than particles came in one of the founding papers of quantum mechanics itself. In 1926, Born, Heisenberg, and Jordan[14] turned their attention to the electromagnetic field in empty space, in the absence of any electric charges or currents. Their work can best be understood by an analogy with Planck's 1900 theory of thermal radiation.

Planck, it will be recalled, had treated the motion of the electrons in a heated body in terms of an idealized picture, in which the electrons were replaced with an unlimited number of modes of simple oscillation, like a violin with a huge number of strings of all possible lengths. He had further proposed that the allowed energies of any one mode of oscillation were separated by a definite quantity, equal to the frequency of the mode times Planck's constant. One of the products of the new quantum mechanics of 1925–1926 was a confirmation of Planck's proposal: it was proved that the energies of a simple oscillator, like a violin string, were indeed quantized, in just the way that Planck had guessed. The essential feature of the dynamics of simple oscillations, used in obtaining this result, is that the energy required to produce any given displacement in the oscillator is proportional to the square of the displacement—as we pull a violin string farther and farther from its equilibrium position, it becomes harder and harder to produce any further displacement.

But essentially the same is true of an electromagnetic field—the energy in any one mode of oscillation of the field is proportional to the square of the field strength—in a sense, to the square of its "displacement" from the normal state of field-free empty space. Thus, by applying to the electromagnetic field the same mathematical methods that they had used for material oscillators, Born et al. were able to show that the energy of each mode of oscillation of an electromagnetic field is quantized—the allowed values are separated by a basic unit of energy, given by the frequency of the mode times Planck's constant. The physical interpretation of this result was immediate. The state of lowest energy is radiation-free empty space, and can be assigned an energy equal to zero. The

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

next lowest state must then have an energy equal to the frequency times Planck's constant, and can be interpreted as the state of a single photon with that energy. The next state would have an energy twice as great, and therefore would be interpreted as containing two photons of the same energy. And so on. Thus, the application of quantum mechanics to the electromagnetic field had at last put Einstein's idea of the photon on a firm mathematical foundation.

Born, Heisenberg, and Jordan had dealt only with the electromagnetic field in empty space, so although their work was illuminating, it did not lead to any important quantitative predictions. The first "practical" use of quantum field theory was made in a 1927 paper of Paul Adrian Maurice Dirac.[15] Dirac was grappling with an old problem: how to calculate the rate at which atoms in excited states would emit electromagnetic radiation and drop into states of lower energy. The difficulty was not so much in deriving an answer—the correct formula had already been derived in an ad hoc sort of way by Born and Jordan[16] and by Dirac[17] himself. The problem was to understand this guessed-at formula as a mathematical consequence of quantum mechanics. This problem was of crucial importance, because the process of spontaneous emission of radiation is one in which "particles" are actually created. Before the event, the system consists of an excited atom, whereas after the event, it consists of an atom in a state of lower energy, plus one photon. If quantum mechanics could not deal with processes of creation and destruction, it could not be an all-embracing physical theory.

The quantum-mechanical theory of such processes can best be understood by returning to the analogy between fields and oscillators. In the absence of any interaction with atoms, the electromagnetic field is like an ensemble of completely isolated violin strings; whatever energy is given to any mode of oscillation, or equivalently, whatever the number of photons of a particular frequency, it will stay the same forever. Similarly, if an atom did not interact with radiation, it would remain indefinitely in whatever state it was placed. But atoms do interact with radiation, because electrons carry an electric charge. So the true analogy is with a set of violin strings that are weakly coupled together, as by the violin soundboard. Every musician knows what happens when one oscillator is set going—it will gradually feed energy into the other modes of oscillation until they are all excited. In quantum mechanics this cannot happen gradually because the energies are quantized, so instead the probability gradually increases that energy which was originally stored in the atom will be found in the electromagnetic field—in other words, that a photon will have been created.

Dirac's successful treatment of the spontaneous emission of radiation confirmed the universal character of quantum mechanics. However, the world was still conceived to be composed of two very different ingredients—particles and fields—which were both to be described in terms of quantum mechanics, but in very different ways. Material particles like electrons and protons were conceived to be eternal; to describe the physical state of a system, one had to describe the probabilities for finding each particle in any given region of space or range of velocities. On the other hand, photons were supposed to be merely a manifestation of an underlying entity, the quantized electromagnetic field, and could be freely created and destroyed.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

It was not long before a way was found out of this distasteful dualism, toward a truly unified view of nature. The essential steps were taken in a 1928 paper of Jordan and Eugene Wigner,[18] and then in a pair of long papers in 1929–1930 by Heisenberg and Pauli.[19] (A somewhat different approach was also developed in 1929 by Enrico Fermi.[20]) They showed that material particles could be understood as the quanta of various fields, in just the same way that the photon is the quantum of the electromagnetic field. There was supposed to be one field for each type of elementary particle. Thus, the inhabitants of the universe were conceived to be a set of fields—an electron field, a proton field, an electromagnetic field—and particles were reduced in status to mere epiphenomena. In its essentials, this point of view has survived to the present day, and forms the central dogma of quantum field theory: the essential reality is a set of fields, subject to the rules of special relativity and quantum mechanics; all else is derived as a consequence of the quantum dynamics of these fields.

This field-theoretic approach to matter had an immediate implication: given enough energy, it ought to be possible to create material particles, just as photons are created when an atom loses energy. In 1932 Fermi[21] used this aspect of quantum field theory to formulate a theory of the process of nuclear beta decay. Ever since Becquerel discovered in 1896 that a crystal containing uranium salts would fog a photographic plate, it was known that nuclei were subject to various kinds of radioactive decay. In one of these modes of decay, known as beta decay, the nucleus emits an electron, and changes its own chemical properties. Throughout the 1920s it was believed that the nuclei are composed of protons and electrons, so there was no great paradox in supposing that every once in a while one of the electrons gets out. However, in 1931 Paul Ehrenfest and Julius Robert Oppenheimer[22] presented a compelling though indirect argument that nuclei do not in fact contain electrons, and in 1932 Heisenberg[23] proposed instead that nuclei consist of protons and the newly discovered neutral particles, the neutrons. The mystery was, where did the electron come from when a nucleus suffered a beta decay? Fermi's answer was that the electron comes from much the same place as the photon in the radiative decay of an excited atom—it is *created* in the act of decay, through an interaction of the field of the electron with the fields of the proton, the neutron, and a hypothesized particle, the neutrino.

One problem remained to be solved after 1930, in order for quantum field theory to take its modern form. In formulating the pre-field-theoretic theory of individual electrons, Dirac[24] in 1928 had discovered that his equations had solutions corresponding to electron states of negative energy, that is, with energy less than the zero energy of empty space. In order to explain why ordinary electrons do not fall down into these negative-energy states, he was led in 1930 to propose[25] that almost all these states are already filled. The unfilled states, or "holes" in the sea of negative energy electrons would behave like particles of positive energy, just like ordinary electrons but with opposite electrical charge: plus instead of minus. Dirac thought at first that these "antiparticles" were the protons, but their true nature as a new kind of particle was revealed with the discovery[26] of the positron in cosmic rays in 1932.

Dirac's theory of antimatter allowed for a kind of creation and annihilation of particles even without introducing the ideas of quantum field theory. Given

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

enough energy, a negative-energy electron can be lifted up into a positive-energy state, corresponding to the creation of a positron (the hole in the negative-energy sea) and an ordinary electron. And of course the reverse annihilation process can also occur. Dirac himself had always resisted the idea that quantum field theory is needed to describe any sort of particle but photons. However, in 1934 a pair of papers by Wendell Furry and Oppenheimer[27] and by Pauli and Victor Weisskopf[28] showed how quantum field theory naturally incorporates the idea of antimatter, without introducing unobserved particles of negative energy, and satisfactorily describes the creation and annihilation of particles and antiparticles. For most theorists, this settled the matter, and particles and antiparticles are now seen as coequal quanta of the various quantum fields.

It is important to understand that quantum field theory gave rise to a new view not only of particles but also of the forces among them. We can think of two charged particles interacting at a distance not by creating classical electromagnetic fields which act on one another, but by exchanging photons, which continually pass from one particle to the other. Similarly, other kinds of force can be produced by exchanging other kinds of particle. These exchanged particles are called *virtual particles*, and are not directly observable while they are being exchanged, because their creation as real particles (e.g., a free electron turning into a photon and an electron) would violate the law of conservation of energy. However, the quantum-mechanical uncertainty principle dictates that the energy of a system that survives for only a short time must be correspondingly highly uncertain, so these virtual particles can be created in intermediate states of physical processes, but must be reabsorbed again very quickly.

From this line of reasoning, one can infer that the force produced by the exchange of a given type of particle has a range (the distance beyond which it falls off very rapidly) inversely proportional to the mass of the exchanged particle. Thus, the photon, which has zero mass, gives rise to a force of infinite range, the familiar inverse-square force of Coulomb. The force between protons and neutrons in an atomic nucleus was known to have a range of a little less than a million millionth of a centimeter, so Hidekei Yukawa[29] was able in 1936 to predict the existence of an entirely new kind of particle, the meson, with a mass a few hundred times that of the electron. In calculating these forces, one assumes that the energy density at a point is not just a sum of squares of fields, as for uncoupled simple mechanical oscillators, but also contains products of the values of the different fields (and their rates of change) at that point. These multifield interactions are the unknowns that have to be sought by our theoretical and experimental efforts. From the viewpoint of quantum field theory, all questions about the particles of which matter is composed and the forces that act among them are only means to an end—the real problem is to determine what are the fundamental quantum fields, and what are the interactions among them.

The Problem of Infinities

I have described the early days of quantum field theory as if it were a grand progress from triumph to triumph. This has been a somewhat distorted picture, for almost from the beginning the theory was thought to be subject to a grave internal inconsistency.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

The problem first appeared in a 1930 paper of Oppenheimer,[30] who was trying to calculate the effect on the energy of an atomic electron produced by its interaction with the quantum electromagnetic field. Just as the exchange of virtual photons between two electrons produces an energy of interaction between them, so also in quantum field theory the emission of virtual photons and their reabsorption by the same electron produces a self-energy, which might depend on the atomic orbit occupied by the electron, and which might show up as an observable shift in atomic energy levels. Unfortunately, Oppenheimer discovered that the energy shift predicted by the quantum theory of the electromagnetic field was infinite!

The infinity here arises because when an electron in an atom turns briefly into a photon and an electron, these two particles can share the momentum of the original electron in an infinite variety of ways. The self-energy of the electron involves a sum over all the ways that the momentum can be shared out, and because there is no limit to how large the momenta can be, this sum turns out to be infinite. It was not obvious that this would have to happen; after all, there are many examples of mathematical series in which one adds up an infinite number of terms and gets a finite result. (For example, $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots$). However, Oppenheimer found that the self-energy of the electron behaved more like the series $1 + 1 + 1 + 1 + \ldots$, and could hardly be interpreted as a finite quantity.

The problem was ameliorated a bit a few years later, when Weisskopf[31] included the effects of processes in which a virtual electron, positron, and photon are created out of empty space, with the positron and photon then being annihilated along with the original electron, leaving the new electron over as a real particle in the final state. This contribution to the self-energy cancelled the worst part of the original infinity found by Oppenheimer, but the self-energy was left in the form of a sum over virtual momenta which behaved like the series $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$, and which still could not be interpreted as a finite quantity.

Similar infinities were found in other problems, such as the "polarization of the vacuum" by applied electric fields,[32] and the scattering of electrons by the electric fields of atoms.[33] (One of the few bright spots in this otherwise discouraging picture came in the treatment of infinities associated with photons of very low momenta; it was shown in 1937 by Felix Bloch and Arnold Nordsieck[34] that these infrared infinities all cancel in the total rates for collisions.) Of course, if one uses a theory to calculate an observable quantity, and finds that the answer is infinite, one concludes either that a mathematical mistake has been made, or that the original theory was no good. Throughout the 1930s, the accepted wisdom was that quantum field theory was in fact no good, that it might be useful here and there as a stopgap, but that something radically new would have to be added in order for it to make sense.

The problem of infinities in fact has provided the single greatest impetus toward a radical revision of quantum field theory. Some of the ideas which were tried out in the 1930s and 1940s are listed below:

1. In 1938 Heisenberg[35] proposed that there is a fundamental unit of energy, and that quantum field theory works only at scales of energy that are small compared with this energy unit. The analogy was with the other fundamental constants, Planck's constant and the speed of light. Quantum mechanics comes

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

into play when ratios of energies and frequencies approaches values as small as Planck's constant, whereas special relativity is needed when velocities approach values as large as the speed of light. In the same way, Heisenberg supposed, some entirely new physical theory might be needed when energies exceed the fundamental unit, and some mechanism in this theory might wipe out the contributions of virtual particles with such high energies, thus avoiding the problem of infinities. (Heisenberg's idea drew support in the 1930s from the observation that the showers of charged particles produced by high energy cosmic rays did not behave as expected in quantum electrodynamics. This discrepancy was later realized to be due to the production of new particles, the mesons, and not to a failure of quantum field theory.)

2. John Archibald Wheeler[36] in 1937 and Heisenberg in 1943[37] independently proposed a positivistic approach to physics, which would have replaced quantum field theory with a different sort of theory, sometimes called an "S-matrix theory," which would involve only directly measurable quantities. They reasoned that experiments do not actually allow us to follow what happens to electrons in atoms or in collision processes. Instead, it is only really possible to measure the energies and a few other properties of bound systems like atoms, and the probabilities for various collision processes. These quantities obey certain very general principles, such as reality, conservation of probabilities, smooth energy dependence, conservation laws, etc., and it was these general principles that were supposed to replace the assumptions of quantum field theory.

3. Dirac[38] in 1942 suggested that quantum mechanics ought to be expanded to include states of negative probability, which could not appear as the initial or final state of any physical process, but which would have to be included among the intermediate states in these processes. In this manner, minus signs might be introduced into the sums over the ways that the intermediate states share out the momentum of the system, so that a finite answer would be obtained, just as $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$ is a finite quantity (the natural logarithm of 2) whereas $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$ is infinite.

4. As already mentioned, Richard Feynman and John Wheeler[39] in 1945 considered the possibility of abandoning field theory altogether, replacing the field-mediated interaction among particles with a direct action at a distance.

Some of these ideas have survived and are now part of the regular equipment of theoretical physics. In particular, the idea of a pure S-matrix theory has flourished in the development of so-called "dispersion relations" which involve only observable quantities.[40] Also, states of negative probability are now a handy mathematical device, useful especially for dealing with the polarization of virtual photons.[41] However, none of these ideas proved to be the key to the problem of infinities.

The solution turned out to be far less revolutionary than most theorists had expected. Recall that the energy of an electron had been found by Oppenheimer, Waller, and Weisskopf to receive an infinite contribution from the emission and reabsorption of "virtual" photons. An infinite self-energy of this sort appears not only when the electron is moving in orbit in an atom, but also when it is at rest in empty space. But special relativity tells us that the energy of a particle at rest is related to its mass by the famous formula $E = mc^2$. Thus the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

electron mass found in tables of physical data could not be just the "bare" mass, the quantity appearing in our equations for the electron field, but would have to be identified with the bare mass *plus* the infinite "self mass," produced by the interaction of the electron with its own virtual photon cloud. This suggests that the bare mass might itself be infinite, with an infinity which just cancels the infinity in the self mass, leaving a finite total mass to be identified with the mass that is actually observed. Of course, it goes against the grain to suppose that a quantity like the bare mass, which appears in our fundamental field equations could be infinite; but after all, we can never turn off the electron's virtual photon cloud to measure the bare mass, so no paradox arises. Similar remarks apply to other physical parameters, like the charge of the electron. Is it then possible that the infinities in the bare masses and bare charges cancel the infinities found in quantum field theory, not only when we calculate the total masses and charges, but in all other calculations as well? This method, of eliminating infinities by absorbing them into a redefinition of physical parameters, has come to be called *renormalization*.

The renormalization idea was suggested by Weisskopf in 1936,[42] and again by Kramers[43] in the mid 1940s. However, it was far from obvious that the idea would work. In order to eliminate infinities by absorbing them into redefinitions of physical parameters, the infinities must appear in only a special way, as corrections to the observed values of these parameters. For instance, in order to absorb the infinite shift in atomic energy levels found by Oppenheimer into a redefinition of the electron mass, the infinite part of the self-energy would have to be the same for all atomic energy levels. The mathematical methods available in the 1930s and early 1940s were simply inadequate for the task of sorting out all the infinities that might appear in all possible calculations to see if they could all be eliminated by renormalization. Perhaps even more important, there was no compelling reason to do so—there were no experimental data which forced theorists to come to grips with these problems. And of course, physicists had other things on their minds from 1939 to 1945.

Revival at Shelter Island

On June 1, 1947, a four-day conference on the foundations of quantum mechanics opened at Shelter Island, a small island near the end of Long Island in New York State. The conference brought together young American physicists from the new generation who had started their scientific work during the war at Los Alamos and the MIT Radiation Laboratory, as well as older physicists who had been active in the 1930s. Among the younger participants was Willis Lamb, an experimentalist then working in the remarkable group of physicists founded at Columbia University by I. I. Rabi. Lamb announced the results of a beautiful experiment, in which he and a student, R. C. Retherford, had for the first time measured an effect of the self-energy of the electron in the hydrogen atom.[44]

The existing theory of the hydrogen atom had been first advanced by Niels Bohr in 1913,[12] then put on a sound mathematical foundation by the quantum mechanics of 1925–1926,[13] and finally corrected to include effects of relativity and the spin of the electron by Heisenberg and Jordan,[45] C. G. Darwin,[46] and

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Dirac.[24] In the final version of this theory, in particular as formulated by Dirac, certain pairs of excited states of the atom were expected to have exactly equal energy. (These pairs of states correspond to the two different ways that the spin of the electron and the angular momentum associated with its revolution around the nucleus could combine to give a definite total angular momentum.) But this theory ignored all effects of the interaction of the electron with its own electromagnetic field—the effects that Oppenheimer[30] had tried to calculate when he discovered the infinities. If such effects were real they would presumably shift the energies of these pairs of states, so that they would no longer be exactly equal.

This is what Lamb and Retherford found. By using the new techniques of handling microwave radiation that had come out of wartime work in radar, they were able to show that the energies of the first two excited states of hydrogen (the $2s_{1/2}$ and $2p_{1/2}$ states), which were supposed to be equal according to the 1928 Dirac theory, actually differed by about 0.4 parts per million. This is now known as the Lamb shift.

Stimulated in part by Lamb's results, the participants at Shelter Island entered into an intense discussion of the underlying theory. I was not at Shelter Island (having just entered high school) and I cannot trace the historical development of the different reformulations of quantum field theory that developed around that time. It would be most valuable for a historian of science to gather the recollections of the participants at Shelter Island and succeeding conferences, read the papers that were written at that time, and put together a coherent account. I will sketch here only a few of the products of this period.

The Lamb shift itself was first calculated by Hans Bethe,[47] I believe on the train ride back from Shelter Island. Using mass renormalization to eliminate infinities, he obtained a result in reasonable agreement with the value announced by Lamb. However, as acknowledged by Bethe himself, this was a rough calculation, involving approximations that were not fully consistent with the Special Theory of Relativity.

There were at least three general reformulations of quantum field theory worked out in the late 1940s that were thoroughly relativistic and that were sufficiently simple and elegant to allow a systematic treatment of the infinities. One of these approaches had actually been developed well before the Shelter Island Conference, by Sin-Itiro Tomonaga[48] and his colleagues in Japan, but I believe that their work had not yet become known in the United States in the summer of 1947. The two other approaches were contributed by participants at Shelter Island, Julian Schwinger[49] and Richard Feynman.[50]

Feynman's work led to a set of pictorial rules which allowed one to associate a definite numerical quantity to each picture of how the momentum and energy could flow through the intermediate states of any collision process: the probability for the process is given by the square of the sum of these individual quantities. The Feynman rules were very much more than a handy calculational algorithm, because they incorporated an essential feature of quantum field theory—the symmetry between particles and antiparticles. Each line in a Feynman diagram can represent either a particle created at one end of the line and destroyed at the other, or an antiparticle going the other way. It is this equal treatment of particles and antiparticles that ensures that the quantities calcu-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

lated by Feynman diagrams are independent of the velocity of the observer, as required by the Special Theory of Relativity, at every stage of the calculation. As had been shown long before by Weisskopf,[31] intermediate states involving antiparticles play a crucial role in cutting down the degree of infinity, from disasters like $1 + 2 + 3 + \ldots$, to something more manageable like $1 + \frac{1}{2} + \frac{1}{3} \ldots$. The Feynman rules automatically ensured the cancellations of the worst infinities, leaving over the more manageable infinities, which could be eliminated by renormalization.

Before the end of 1947 Schwinger[51] had used his approach to carry out what I believe was the first calculation of another effect of the electron's cloud of virtual photons, the *anomalous magnetic moment* of the electron. One of the triumphs of Dirac's 1928 theory[24] was its prediction of the magnetic moment of the electron, a number which characterizes the strength of the electron's interaction with magnetic fields, and the strength of its own magnetic field. However, experiments[52] at Columbia in 1947 had revealed that the magnetic moment of the electron is actually a little larger than the Dirac value, by 1.15 to 1.21 parts per thousand. By absorbing the infinite effects of virtual photons into a renormalization of the charge of the electron, Schwinger was able to calculate a finite magnetic moment that was larger than the Dirac value by just 1.16 parts per thousand!

Of course, both the experimental and the theoretical determinations of effects like the Lamb shift and the anomalous magnetic moment of the electron have been enormously improved since 1947.[53] For instance, right now the experimental value of the magnetic moment of the electron is larger than the Dirac value by 1.15965241 parts per thousand, whereas the theory gives this anomalous magnetic moment as 1.15965234 parts per thousand, with uncertainties of about 0.00000020 and 0.00000031 parts per thousand, respectively. The precision of the agreement between theory and experiment here can only be called spectacular.

Finally, Freeman Dyson[54] in 1949 showed that the formalisms of Schwinger and Tomonaga would yield the same graphical rules that had been found by Feynman. Dyson also carried out an analysis of the infinities in general Feynman diagrams, and sketched out a general proof that these infinities are always of precisely the sort which could be removed by renormalization.[55] As a graduate student in the mid-1950s, I learned the new approach to quantum field theory by reading Dyson's marvelously lucid papers.

I should emphasize that the theory of Schwinger, Tomonaga, Feynman, and Dyson was not really a new physical theory. It was simply the old quantum field theory of Heisenberg, Pauli, Fermi, Oppenheimer, Furry, and Weisskopf, but cast in a form far more convenient for calculation, and equipped with a more realistic definition of physical parameters like masses and charges. The continued vitality of the old quantum field theory after fifteen years of attempts to find a substitute is truly impressive.

This raises an interesting historical question. All the effects that were calculated in the great days of 1947–1949 could have been estimated if not actually calculated at any time after 1934. True, without the renormalization idea, the answers would have been formally infinite, but at least it would have been possible to guess the order of magnitude of quantities like the Lamb shift and

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

the correction to the magnetic moment of the electron. (An infinite series like $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$ grows very slowly; after a million terms, it is still less than 14.4.) Not only was this not done—most theorists seemed to have believed that these quantities were zero! Indeed, some evidence for what was later called the Lamb shift had actually been discovered[56] in 1938, but to the best of my knowledge, no theorist checked to see whether the order of magnitude of this reported energy splitting was more or less what would be expected in a quantum field theory.

Why was quantum field theory not taken more seriously? One reason is the tremendous prestige of the Dirac theory[24] of 1928, which had worked so well in accounting for the fine structure of the hydrogen spectrum without including self-energy effects. Even more important, the appearance of infinities discredited quantum field theory altogether in many physicists' minds. But I think that the deepest reason is a psychological difficulty, that may not have been sufficiently appreciated by historians of science.[57] There is a huge apparent distance between the equations that theorists play with at their desks, and the practical reality of atomic spectra and collision processes. It takes a certain courage to bridge this gap, and to realize that the products of thought and mathematics may actually have something to do with the real world. Of course, when a branch of science is well under way, there is continual give and take between theory and experiment, and one gets used to the idea that the theory is about something real. Without the pressure of experimental data, the realization comes harder. The great thing accomplished by the discovery of the Lamb shift was not so much that it forced us to change our physical theories, as that it forced us to take them seriously.

Weak and Strong Interactions

For a few years after 1949, enthusiasm for quantum field theory was at a high level. Many theorists expected that it would soon lead to an understanding of all microscopic phenomena, not only the dynamics of photons, electrons, and positrons. However, it was not long before there was another collapse in confidence—shares in quantum field theory tumbled on the physics bourse, and there began a second depression, which was to last for almost twenty years.

Part of the problem arose from the limited applicability of the renormalization idea. In order for all infinities to be eliminated by a renormalization of physical parameters like masses and charges, it is necessary for these infinities to arise in only a limited number of ways, as corrections to masses, charges, etc., but not otherwise. Dyson's work[54] showed that this would be the case for only a small class of quantum field theories, which are called *renormalizable* theories. The simplest theory of photons, electrons, and positrons (known as *quantum electrodynamics*) is renormalizable in this sense, but most theories are not.

Unfortunately, there was one important class of physical phenomena which apparently could not be described by a renormalizable field theory. These were the *weak interactions*, which cause the radioactive beta decay of nuclei mentioned above in the section on the Birth of Quantum Field Theory. Fermi[21] had invented a theory of weak interactions in 1932 which, with a few modifications,

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

adequately described all weak interaction phenomena in the lowest order of approximation, that is, including only a single simple Feynman diagram in calculations of transition rates. However, as soon as this theory was pushed to the next order of approximation, it exhibited infinities which could not be removed by a redefinition of physical quantities.

The other major problem had to do with the limited validity of the approximation techniques used in 1947–1949. Any physical process is represented by an infinite sum of Feynman diagrams, each one representing a particular sequence of intermediate states consisting of definite numbers of particles of various types. To each diagram we associate a numerical quantity; the rate of the process is the square of the sum of these quantities. Now, in quantum electrodynamics the quantities associated with complicated diagrams are very small; for each additional photon line there is one additional factor of a small number known as the *fine-structure constant*, roughly 1/137. In the Fermi theory of weak interactions, the corresponding factor is even smaller—at the typical energies of elementary particle physics, it is 10^{-5} to 10^{-7}. It is the rapid decrease of the contributions associated with complicated Feynman diagrams (together with the renormalizability of the theory) that makes it possible to carry calculations in quantum electrodynamics to such a fantastic degree of accuracy.

However, in addition to electromagnetic and weak interactions, there is one other class of interaction in elementary particle physics, known as the *strong interactions*. It is the strong interactions that hold atomic nuclei together against the electrostatic repulsion of the protons they contain. For strong interactions, the factor corresponding to the fine structure constant is roughly of the order of one instead of 1/137, so complicated Feynman diagrams are just as important as simple ones. (This of course is why these interactions are strong.) Thus, although attempts have been made time and time again to use quantum field theory in calculations of the nuclear force, it has never really worked in a convincingly quantitative way. New kinds of strongly interacting particles called mesons and hyperons were being discovered from 1947 on, first in cosmic rays and then in accelerator laboratories, and quantum field theory was at first enthusiastically used to study their strong interactions, but again with little quantitative success. It was not that there was any difficulty in thinking of renormalizable quantum field theories that *might* account for the strong interactions—it was just that having thought of such a theory, there was no way to use it to derive reliable quantitative predictions, and to test if it were true.

The nonrenormalizability of the field theory of weak interactions and the uselessness of the field theory of strong interactions led in the early 1950s to a widespread disenchantment with quantum field theory. Some theorists turned to the study of symmetry principles and conservation laws, which can be applied to physical phenomena without detailed dynamical calculations. Others picked up the old S-matrix theory of Wheeler and Heisenberg, and worked to develop principles of strong interaction physics that would involve only observable quantities. In both lines of work, quantum field theory was used heuristically, as a guide to general principles, but not as a basis for quantitative calculation.

Now, as a result of work by many physicists over the last decade, quantum

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

field theory has again become what it was in the late 1940s, the chief tool for a detailed understanding of elementary particle processes. There are quantum field theories of the weak and strong interactions, called *gauge theories*, which are not subject to the old problems of nonrenormalizability and incalculability, and which are to some extent even true! We are still in the midst of this revival, and I will not try to outline its history, but will only summarize how the old problems of quantum field theory are surmounted in the new theories.[58]

The essence of the new theories is that the weak and the strong interactions are described in a way that is almost identical to the successful older quantum field theory of electromagnetic interactions. Just as electromagnetic interactions among charged particles are produced by the exchange of photons, so the weak interactions are produced by the exchange of particles called *intermediate vector bosons* and the strong interactions by the exchange of other particles called *gluons*. All these particles—photons, intermediate vector bosons, and gluons—have equal spin, and have interactions governed by certain powerful symmetry principles known as *gauge symmetries*. (A gauge symmetry principle states that the fundamental equations do not change their form when the fields are subjected to certain transformations, whose effect varies with position and time.) Because these theories are so similar to quantum electrodynamics, they share its fundamental property, of being renormalizable. Indeed, the relation between weak and electromagnetic interactions is not merely one of analogy—the theory unifies the two, and treats the fields of the photon and the intermediate vector bosons as members of a single family of fields.

The intermediate vector bosons are not massless, like the photon, but instead are believed to have perhaps 70 to 80 times the mass of a proton or neutron. This huge mass is not due to any essential dissimilarity between the photon and the intermediate vector boson fields, but instead arises from the way that the symmetry of the underlying field theory breaks down when the field equations are solved. The family of intermediate vector bosons, of which the photon is a member, is believed to contain one heavy charged particle and its antiparticle, called the W^+ and W^-, and one even heavier neutral particle, called the Z^0. Exchange of the W produces the familiar weak interactions, like nuclear beta decay, whereas exchange of the Z^0 would produce a new kind of weak interaction, in which the participating particles do not change their charge. Such *neutral current processes* were discovered in 1973, and are found to have just about the properties expected in these theories. All of the intermediate vector bosons are much too heavy to have been produced with existing accelerator facilities, but there are great hopes of producing them with colliding beams of protons and antiprotons before too long.

In contrast, the gluons which mediate the strong interactions may well have zero mass. Such theories with massless gluons have a remarkable property known as *asymptotic freedom*—at very high energy or very short distances, the strength of the gluon interactions gradually decreases. In consequence, it is now possible to use quantum field theory to carry out detailed calculations of strong interaction processes at sufficiently high energy. In particular, it has been possible to account for some of the features observed in a process such as high energy electron-nucleon scattering.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Just as the gluon interactions become weak at high energy and short distances, they also become strong at low energy and long distances. For this reason, it is widely believed (though not yet proved) that particles which carry the quantity called *color* with which gluons interact (in the same sense that photons interact with electric charge) cannot be produced as separate free particles. The colored particles include the gluons themselves, which is presumably why gluons have never been observed as real particles. The colored particles are also believed to include the quarks discussed in this double issue of *Daedalus* in the article by Sidney Drell. The observed strongly interacting particles such as neutrons, protons, and mesons are believed to be compound states, consisting of quarks, antiquarks, and gluons, but with no net color. This picture represents a nearly complete triumph of the field over the particle view of matter: the fundamental entities are the quark and gluon fields, which do not correspond to any particles that can be observed even in principle, whereas the observed strongly interacting particles are not elementary at all, but are mere consequences of an underlying quantum field theory.

There are hopes of a unified gauge theory of weak, electromagnetic, and strong interactions. The photon, intermediate vector bosons, and gluons would then form part of a single family of fields. However, in order for this to be possible, there would have to be other fields in this family, corresponding to particles of extraordinarily high mass. According to one estimate, the expected mass of these new particles would be 10^{17} (a hundred thousand million million) proton masses. These masses are so high that it is no longer possible to ignore the gravitational fields of these particles, as done almost everywhere else in particle physics.

Unfortunately, despite strenuous efforts which continue to the present, there still has not been found a satisfactory (e.g., renormalizable) quantum field theory of gravitation. It is ironic that gravitation, which provided the first classical field theory, has so far resisted incorporation into the general framework of quantum field theory.

Throughout this history I have put great emphasis on the condition of renormalizability, the requirement that it should be possible to eliminate all infinities in a quantum field theory by a redefinition of a small number of physical parameters. Many physicists would disagree with this emphasis, and indeed, it may eventually be found that all quantum field theories, renormalizable or not, are equally satisfactory. However, it has always seemed to me that the requirement of renormalizability has just the kind of restrictiveness that we need in a fundamental physical theory. There are very few renormalizable quantum field theories. For instance, it is possible to construct quantum field theories of electromagnetism in which the electron has any magnetic moment we like, but only one of these theories, corresponding to a magnetic moment of 1.0011596523 . . . times the Dirac value, is renormalizable. Also, as we have seen, it took a long time before it was found that there are any renormalizable theories at all of the weak interactions. We very much need a guiding principle like renormalizability to help us to pick the quantum field theory of the real world out of the infinite variety of conceivable quantum field theories. Thus, if renormalizability is ultimately to be replaced with some other condition, I

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

would hope that it will be one that is equally or even more restrictive. After all, we do not want merely to describe the world as we find it, but to explain to the greatest possible extent why it has to be the way it is.

REFERENCES

I am grateful for valuable advice from Felix Bloch, Freeman Dyson, Francis Everitt, I. I. Rabi, and Victor Weisskopf.

[1]This often quoted summary of Newton's theory of gravitation appears at the end of the *Philosophiae Naturalis Principia Mathematica*, p. 546, in the translation of A. Motte, revised and annotated by F. Cajori (Berkeley: University of California Press, 1966), p. 546.

[2]J. C. Maxwell, article on "Ether" in *The Encyclopedia Britannica*, 9th ed. (1875–1889); reprinted in W. D. Niven (ed.), *The Scientific Papers of James Clerk Maxwell* (New York: Dover, 1965), p. 763. Also see Maxwell's *Treatise on Electricity and Magnetism*, vol. 2 (New York: Dover, 1954), pp. 492-493. A useful short account of Maxwell's life and work is given by C. W. F. Everitt, *James Clerk Maxwell—Physicist and Natural Philosopher* (New York: Scribner, 1974).

[3]See Secs. 600-601, 668, and 769 of Maxwell, *Treatise*.

[4]W. Wien, *Ann. Phys.*, 5 (1901): 501.

[5]M. Abraham, *Nachr. Ges. Wiss.* (Gottingen), (1902): 20; *Ann. Phys.*, 10 (1903): 105; *Phys. Zeit.*, 5 (1904): 576.

[6]H. A. Lorentz, *Phys. Zeit.*, 1 (1900): 498, 514; *Versl. Kon. Akad. Wetenschap.* (Amsterdam), 9 (1900): 418; *Proc. Roy. Acad. Amsterdam*, 6 (1904): 809; *Encylopädie der Math. Wiss.* V/2 (1904), pp. 63, 145.

[7]H. Poincaré, *Rend. Circ. Mat.* (Palermo), 21 (1905): 129.

[8]A. Einstein, *Ann. Phys.*, 17 (1905): 891, 905; 18 (1905): 639.

[9]M. Planck, *Verhand. Deutsch. Phys. Ges.*, 2 (1900): 237; reprinted with an English translation by H. Kangro in *Planck's Original Papers in Quantum Physics* (New York: Wiley, 1972).

[10]G. Gamow, *Thirty Years That Shook Physics* (Garden City, N.Y.: Doubleday, 1966), p. 23.

[11]A. Einstein, *Ann. Phys.*, 17 (1905): 132.

[12]N. Bohr, *Phil. Mag.*, 26 (1913): 1, 476, 857.

[13]L. de Broglie, *Compt. rend.*, 177 (1923): 507, 548, 630; W. Heisenberg, *Z. Phys.*, 33 (1925): 879; M. Born and P. Jordan, *Z. Phys.*, 34 (1925): 858; P. A. M. Dirac, *Proc. Roy. Soc.*, A109 (1925): 642; M. Born, W. Heisenberg, and P. Jordan, *Z. Phys.*, 36 (1926): 336; E. Schrödinger, *Ann. Phys.*, 79: 361, 489; 80: 437; 81: 109 (all 1926). Many of these papers are reprinted in English translations in B. L. van der Waerden, *Sources of Quantum Mechanics* (New York: Dover, 1968).

[14]Born, Heisenberg, and Jordan, *Z. Phys.*, 36 (1926): 336, Sec. 3.

[15]P. A. M. Dirac, *Proc. Roy. Soc.* (London), A114 (1947): 243. This article is reprinted in an invaluable collection of papers on quantum field theory, J. Schwinger (ed.), *Quantum Electrodynamics* (New York: Dover, 1958), referred to below as *QED*.

[16]Born and Jordan, *Z. Phys.*, 36 (1926): 336.

[17]P. A. M. Dirac, *Proc. Roy. Soc.* (London), A112 (1926): 661, Sec. 5.

[18]P. Jordan and E. Wigner, *Z. Phys.*, 47 (1928): 631, reprinted in *QED*.

[19]W. Heisenberg and W. Pauli, *Z. Phys.*, 56 (1929): 1; ibid, 59 (1930): 168.

[20]E. Fermi, *Rend. R. Acc. dei Lincei* (6) 9 (1929): 881.

[21]E. Fermi, *Z. Phys.*, 88 (1934): 161.

[22]P. Ehrenfest and J. Oppenheimer, *Phys. Rev.*, 37 (1931): 333.

[23]W. Heisenberg, *Z. Phys.*, 77 (1932): 1; also see I. Curie-Joliot and F. Joliot, *Compt. rend.*, 194 (1932): 273.

[24]P. A. M. Dirac, *Proc. Roy. Soc.* (London), A117 (1928): 610.

[25]P. A. M. Dirac, *Proc. Roy. Soc.* (London), A126 (1930): 360; *Proc. Camb. Phil. Soc.*, 26 (1930): 361.

[26]C. D. Anderson, *Science*, 76 (1932): 238; *Phys. Rev.*, 43 (1933): 491.

[27]W. H. Furry and J. R. Oppenheimer, *Phys. Rev.*, 45 (1934): 245.

[28]W. Pauli and V. Weisskopf, *Helv. Phys. Acta*, 7 (1934): 709.

[29]H. Yukawa, *Proc. Phys.-Math. Soc.* (Japan), 17 (3) (1935): 48.

[30]J. R. Oppenheimer, *Phys. Rev.*, 35 (1930): 461. Also see I. Waller, *Z. Phys.* 59 (1930): 168; ibid., 62 (1930): 673.

[31]V. F. Weisskopf, *Z. Phys.*, 89 (1934): 27; ibid., 90 (1934): 817. Also, *Phys. Rev.*, 56 (1939): 72, reprinted in *QED*.

[32]P. A. M. Dirac, in *XVII Conseil Solvay de Physique* (Brussels, 1933), p. 203.

[33]S. M. Dancoff, *Phys. Rev.*, 55 (1939): 959.

[34]F. Bloch and A. Nordsieck, *Phys. Rev.* 52 (1937): 54, reprinted in *QED*.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

[35]W. Heisenberg, *Ann. Phys.*, 32 (1938): 20.

[36]J. A. Wheeler, *Phys. Rev.*, 52 (1937): 1107.

[37]W. Heisenberg, *Z. Phys.*, 120 (1943): 513, 673; *Z. Naturforsch.*, 1 (1946): 608. Also see C. Moller, *Kon. Dansk. Vid. Sels. Mat.-Fys. Medd.*, 23 (1) (1945); ibid., (19) (1946). (Heisenberg actually introduced a fundamental length rather than an energy, but the two are equivalent notions: any length determines an energy, the energy of a photon with that wavelength).

[38]P. A. M. Dirac, *Proc. Roy. Soc.* (London), A180 (1942): 1.

[39]J. A. Wheeler and R. P. Feynman, *Rev. Mod. Phys.*, 17 (1945): 157; ibid., 21 (1949): 425.

[40]See e.g. G. F. Chew, *The S-Matrix Theory of Strong Interactions* (New York: Benjamin, 1961).

[41]See e.g. S. N. Gupta, *Proc. Phys. Soc.*, 63 (1950): 681; 64 (1951): 850.

[42]V. F. Weisskopf, *Kon. Dan. Vid. Sel. Mat.-Fys. Medd.*, 15 (6) (1936), esp. p. 34 and pp. 5-6; reprinted in *QED*.

[43]The only published report of Kramers' work that I can find is in the *Rapports du VIII Conseil de Physique Salvay* 1948 (ed. R. Stoops, Brussels, 1950). However, Kramers reported on his work at the Shelter Island Conference.

[44]W. E. Lamb, Jr. and R. C. Retherford, *Phys. Rev.*, 72 (1947): 241, reprinted in *QED*.

[45]W. Heisenberg and P. Jordan, *Z. Phys.*, 37 (1926): 263.

[46]C. G. Darwin, *Proc. Roy. Soc.* (London), A116 (1927): 227.

[47]H. Bethe, *Phys. Rev.* 72 (1947): 339, reprinted in *QED*.

[48]S. Tomonaga, *Prog. Theor. Phys.* (Japan), 1 (1946): 27; Z. Koba, T. Tati, and S. Tomonaga, ibid., 2 (1947): 101; S. Kanesawa and S. Tomonaga, ibid., 3 (1948): 276; Z. Koba and S. Tomonaga, ibid., 3 (1948): 290. Some of these articles are reprinted in *QED*.

[49]J. Schwinger, *Phys. Rev.* 74 (1948): 1439; 75 (1949): 651; 76 (1949): 790; 82 (1951): 664, 914; 91 (1953): 713; *Proc. Nat. Acad. Sci.* 37 (1951): 452. All but the first of these articles are reprinted in *QED*.

[50]R. P. Feynman, *Rev. Mod. Phys.*, 20 (1948): 367; *Phys. Rev.* 74 (1948): 939, 1430; 76 (1949): 749, 769; 80 (1950): 440. Some of these articles are reprinted in *QED*.

[51]J. Schwinger, *Phys. Rev.*, 73 (1948): 146, reprinted in *QED*.

[52]J. E. Nafe, E. B. Nelson, and I. I. Rabi, *Phys. Rev.*, 71 (1947): 914; D. E. Nagel, R. S. Julian, and J. R. Zacharias, *Phys. Rev.*, 72 (1947): 97; P. Kusch and H. M. Foley, *Phys. Rev.*, 72 (1947): 1256.

[53]For early calculations of the Lamb shift, see J. B. French and V. F. Weisskopf, *Phys. Rev.*, 75 (1949): 1240; N. M. Kroll and W. E. Lamb, ibid. (1949): 388; R. P. Feynman, *Phys. Rev.*, 74 (1948): 1430; H. Fukuda, Y. Miyamoto, and S. Tomonaga, *Prog. Theor. Phys.* (Japan), 4 (1948): 47, 121. The article by Kroll and Lamb is reprinted in *QED*.

[54]F. J. Dyson, *Phys. Rev.*, 75 (1949): 486, 1736, reprinted in *QED*.

[55]Elements needed to complete Dyson's proof were later supplied by A. Salam, *Phys. Rev.*, 82 (1951): 217; 84 (1951): 426; and S. Weinberg, *Phys. Rev.* 118 (1960): 838.

[56]S. Pasternack, *Phys. Rev.*, 54 (1938): 1113. Also see W. V. Houston, *Phys. Rev.*, 51 (1937): 446; R. C. Williams, *Phys. Rev.*, 54 (1938): 558. Contrary results were reported by J. W. Drinkwater, O. Richardson, and W. E. Williams, *Proc. Roy. Soc.*, 174 (1940): 164.

[57]I have recently discussed this point in a quite different context in *The First Three Minutes—A Modern View of the Origin of the Universe* (New York: Basic Books, 1977), Ch. 6.

[58]For a review of this subject on a technical level, with references to the original literature, see e.g. S. Weinberg, *Rev. Mod. Phys.*, 46 (1974): 255. I have also given nonmathematical accounts in *Scientific American*, 231 (July 1974): 50; and in *Bull. Am. Acad. Arts and Sciences*, 24 (4) (1976): 13; reprinted (in Spanish) in *Plural*, (1976); and in *American Scientist*, 65 (March 1977): 171. Aspects of this subject are treated in two recent books: J. C. Taylor, *Gauge Theories of Weak Interactions* (Cambridge, England: Cambridge University Press, 1976); G. Feinberg, *What Is the World Made Of? The Achievements of Twentieth Century Physics* (Garden City, N.Y.: Anchor Press/Doubleday, 1977).

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

BRUNO ROSSI

X-Ray Astronomy

The Beginnings

IN THE IMPRESSIVE ADVANCE of natural sciences that has been a characteristic feature of the cultural development of the present century, progress in astrophysics has held an increasingly prominent position. Not only have we achieved a sound interpretation, based on well-established physical laws, of many empirical results of previous observations, but new celestial objects and new celestial phenomena have been discovered, which have drastically changed our view of the Universe.

Several factors have contributed to these developments, among them the very great increase in the range and refinement of optical observations, made possible by the availability of more powerful telescopes and more sophisticated observational techniques; the application to astrophysical phenomena of the new advances in experimental and theoretical physics, particularly atomic and nuclear physics; and, during the last few decades, the technical developments which have enabled astronomers to extend their observations, in both directions, outside the narrow range of wavelengths that comprises the optical spectrum. X-ray astronomy is the most recent result of these technical developments.

X-rays are part of the electromagnetic spectrum. Electromagnetic radiations partake of the properties of both waves and particles; radiations belonging to different parts of the spectrum may be described as waves of different wavelengths or as photons (i.e., light quanta) of different energies. Photon energies are usually measured in electronvolts (ev), kiloelectronvolts (kev = 1000 ev), megaelectronvolts (mev = 1,000,000 ev). In order of decreasing wavelengths, that is, in order of increasing photon energies, the electromagnetic spectrum includes radio waves, infrared rays, visible light, ultraviolet rays, X-rays, and gamma rays. The Earth's atmosphere strongly absorbs most of the electromagnetic spectrum; the exceptions are a narrow portion including mainly visible light (the optical window) and another wider portion, in the range of radio waves (the radio window). Ordinary telescopes look at the sky through the optical window. Radiotelescopes look at the sky through the radio window.

All the rest of the electromagnetic spectrum became available to astronomers only when the development of sounding rockets and artificial satellites made it

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

possible to send their sensing instruments above the atmosphere, with a more limited but still important role played by the advances in the technology of *high-flying balloons*.

Outside the atmosphere nothing interferes with the observation of radiations emanating from our closest celestial neighbors—the Sun and the other objects of the solar system. But for the observation of objects beyond the solar system there is a further obstacle: the absorption by interstellar matter whose extreme dilution, for certain wavelengths at least, is counterbalanced by the enormous distances involved.

Interstellar matter is mainly in the form of a gas, to a large extent hydrogen. This gas is perfectly transparent from radio waves to ultraviolet rays of 13.5 ev photon energy. Above this energy interstellar gas become suddenly very opaque for galactic distances. As the wavelength decreases further, the absorption decreases, the gradual drop of the absorption curve being interrupted occasionally by sudden increases.* Eventually the interstellar absorption becomes again negligible. This happens in the X-ray part of the spectrum, at photon energies ranging from about 100 ev for the closest galactic objects to several thousand ev (several kev) for those farthest away. It follows that, for astronomical purposes, the electromagnetic spectrum is split into two regions: a *low-energy* region (extending from radio waves to the near ultraviolet) and a *high-energy* region (including X-rays and gamma rays), the two regions being separated by a "blind" region caused by interstellar absorption.

The whole of the high-energy region has been actively investigated in recent years. However, it is in the X-ray region, specifically in the range of photon energies extending from somewhat less than 1 kev to about 100 kev, that the most abundant, precise, and detailed observational data have been obtained. (For most medical diagnostic purposes, X-rays have energies around 50 kev.) Indeed X-ray astronomy has reached today a degree of maturity approaching that of optical and radio astronomy.

The purpose of this paper is to present a brief history of X-ray astronomy from its inception to the present day. I am painfully aware that, in limiting subject matter to some of the most significant results, and in attempting not to burden the paper with long lists of names, I shall neglect to give proper credit to many distinguished scientists who have made important contributions to X-ray astronomy, often mentioning a team, rather than its individual members. To them I offer my apologies, trusting in their indulgence.

One may place the beginning of X-ray astronomy in the late 1940s, when a sounding rocket launched by the Naval Research Laboratory (NRL) first detected a weak X-ray flux from the Sun. The following years saw the development of a vigorous solar X-ray program. However, it was not until the early 1960s that the first extrasolar X-ray source was discovered.

Because I had a part in the birth of extrasolar X-ray astronomy, I may

*The absorption curve of interstellar gas has the following explanation: at 13.5 ev photons become capable of splitting (ionizing) hydrogen atoms, and they disappear in this process. As the photon energy increases further the photons become less effective in ionizing hydrogen, and less easily absorbed. The absorption undergoes sudden increases at the photon energies corresponding to the ionization energies of helium and of the less abundant, heavier components of interstellar gas.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

perhaps be allowed to set forth here my personal recollections about this event. In the first place I would like to make it clear that the discovery, in 1962, of the first X-ray source outside the solar system (to be named later Scorpio X–1) was neither the result of a theoretical prediction, based on known facts; nor was it accidental, as one might surmise from the fact that the "official" aim of the rocket flight which detected Sco X–1 was to observe X-rays from the Moon. The initial motivation of the experiment which led to this discovery was instead a subconscious trust of mine in the inexhaustible wealth of nature, a wealth that goes far beyond the imagination of man. This meant that, whenever technical progress opened a new window into the surrounding world, I felt the urge to look through this window, hoping to see something unexpected.

In the present case, the technical progress was the development of space technology, to which, distracted by other activities, I had not paid much attention until the summer of 1958, when, as a member of the Space Science Board of the National Academy of Sciences, I became involved in the planning of scientific experiments in space. Because no one had yet searched the sky from outer space with detectors substantially more sensitive than those used for solar X-ray astronomy, an entirely unexplored territory lay open before us, provided only that such detectors could be built.

Spurred by this thought, in the fall of 1959, I suggested a program in extrasolar X-ray astronomy to American Science and Engineering (AS&E), a private company in Cambridge, Massachusetts, for which I was acting as a consultant. This proposal was enthusiastically accepted by the company's president, Martin Annis, who was himself a scientist with experience in cosmic-ray research. Riccardo Giacconi, also a former cosmic-ray physicist who had recently joined the company, became greatly interested and took charge of the program. Other scientists joined in the discussions, most actively George Clark and Stanislaw Olbert of the Massachusetts Institute of Technology (MIT), who were also active as consultants for AS&E. We soon realized that all astronomical information available at that time led to the conclusion that any conceivable X-ray source located outside the solar system was extremely unlikely to produce a signal strong enough to be detected, except by instruments very far beyond the state of the art. This was clearly the reason why scientists who had previously realized the potential interest of extrasolar X-ray astronomy, had not undertaken an observational program in this field. To complicate matters further, there was the practical consideration that it would have been impossible to find financial support for such an unpromising project as to look for extrasolar X-ray sources before achieving a major advance in the detection techniques.

Out of these discussions, two lines of attack emerged. The first had the long-range aim of developing an entirely new type of instrument, capable of achieving sensitivities not within reach of existing techniques. The basic difficulty was the fact that X-ray detectors are also sensitive to cosmic rays—streams of bare atomic nuclei entering the Earth's atmosphere from outer space. Therefore, building very large detectors did not necessarily provide the best solution, because the very weak X-ray sources would be swamped by the cosmic-ray background.

To overcome this difficulty, Giacconi suggested the development of a device capable of concentrating over a small detector the radiation from a distant X-ray

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

source, incident upon a large surface; and to use for this purpose the fact that X-rays undergo total reflection when striking a smooth surface at a sufficiently shallow angle.

The principle of grazing incidence optics for X-ray microscopy had been set forth by H. Wolter in the early 1950s. In particular Wolter had shown that true X-ray images could be formed by double grazing incidence reflection from properly shaped surfaces. Having become acquainted with this work, Giacconi planned a program for the development of grazing incidence optics aiming ultimately at the construction of a true telescope.

The program got under way in 1960 at AS&E, under NASA sponsorship. It took a great deal of ingenuity and of hard work, extending over a period of years, to actually achieve this aim. Comparatively small, but highly refined X-ray telescopes, flown on rockets and on the occasion of the Skylab mission, have already produced remarkable X-ray pictures of the Sun, making important contributions to the interpretation of solar phenomena. Only recently X-ray optics has begun to be used for extrasolar X-ray astronomy.

The second line of attack was to do the best that could be done with instruments similar to those already in existence. Because of the prevailing skepticism about the possibility of detecting sources outside the solar system, it was decided to propose an experiment aiming at the detection of X-rays from the Moon. (From what we knew at that time, we estimated that the Moon would be the strongest X-ray source in the sky after the Sun, X-rays being produced on the lunar surface by the impact of solar X-rays, or possibly of high-energy electrons, originating from the Sun).

This experiment, for which AS&E succeeded in securing the support of the U.S. Air Force, through its Cambridge Research Laboratories, might have yielded some useful information on the chemical composition of the lunar surface. But for us, whose main interest was extrasolar astronomy, the lunar experiment was attractive because it provided a convenient stepping stone in our more ambitious program. Moreover, we still clung to the hope that, contrary to all reasonable expectations, we might detect extrasolar X-rays.

Most of the detectors used for the study of solar X-rays had been Geiger counters provided with small (about 1 cm^2) very thin windows to admit X-rays into the sensitive volume. Frank Paolini at AS&E developed Geiger counters with thin mica windows of much larger area (about 20 cm^2, a difficult achievement at that time). An anticoincidence device, of a kind well known to cosmic-ray physicists, was employed to reduce the cosmic-ray background. The detectors had also large fields of view. The result of these design features was that the detectors had about one hundred times greater sensitivity than any of those flown previously.

The first successful flight with these detectors was carried out in June, 1962. This experiment, in fact, did not detect X-rays from the Moon; it discovered instead Sco X–1. The discovery was published in late 1962 by Riccardo Giacconi, Herbert Gursky, Frank Paolini, and myself.

The reason why no one had expected any significant result in extrasolar X-ray astronomy from experiments carried out with the rather rudimentary means available in 1962—and the reason why our findings were received at first with great skepticism by much of the astronomical community—was not a failure of

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

the theory (although sometimes theories are proved incorrect.) It was rather the reluctance to consider the possible existence of celestial objects entirely different from those already known. The X-ray emission from the Sun, above 0.5 kev, is never more than one part in one million of the visible emission, even during large solar flares; it is about one part in ten billion in the absence of activity; whereas, as it turned out, Sco X–1 radiates about one thousand times more energy in the form of X-rays than in the form of light.

It took about one and half years, and several additional rocket flights, by the AS&E group and other groups, and the observation of a few other X-ray sources, before the early misgivings about the credibility of our discovery were dispelled. Particularly important, in this respect, was a rocket flight carried out in April 1963, by the NRL group headed by Herbert Friedman, which confirmed the existence of the source discovered by the AS&E group and pinpointed its position in the sky more precisely, placing it in the constellation of Scorpio; hence the name of Sco X–1 by which this source became known.

There immediately arose a number of problems; foremost among them was the identification of the visible celestial objects responsible for the X-ray emission. It was quite natural to think that, in ordinary light, these objects would appear quite conspicuous, easily distinguishable from ordinary stars, and this expectation was soon partially fulfilled by the identification with the Crab Nebula of one of the sources seen in the early experiments.

In the late twenties, it had been established that the Crab Nebula (a gas cloud, a few arc minutes in diameter†) is the remnant of a supernova, that is, of a cataclysmic stellar explosion, observed, according to the Chinese and Japanese records, in the year 1054. It is one of the most interesting galactic objects, a kind of celestial laboratory in which a great many remarkable phenomena take place. For this reason, during the early discussions on the aims of X-ray astronomy, the Crab Nebula had been singled out as one of the most likely sources of celestial X-rays, even though estimates of its X-ray intensity fell considerably short of that which was later actually observed.

The identification—a milestone in the history of X-ray astronomy—was the result of an experiment performed in the summer of 1964 by the NRL group under the leadership of Friedman. Having succeeded in launching an X-ray rocket when the Crab Nebula was about to be eclipsed by the Moon, they observed the simultaneous disappearance of the X-ray source and of the visible nebula.

The identification of the other X-ray sources proved much more elusive. Finally, in 1966, it was established that the visible counterpart of Sco X–1 was an inconspicuous, faint (13th magnitude) star which, until that time, had not attracted the attention of optical astronomers. This discovery—still another milestone in our history—was the result of a very precise determination of the position of Sco X–1 by a team of scientists from AS&E and MIT, followed by optical observations carried out at the Observatory of Tokyo and the Palomar Observatory. It was the first example of the essential role to be played in the

†Angles are measured in degrees, arc minutes, and arc seconds. One degree includes 60 arc minutes; one arc minute 60 arc seconds. The Moon and the Sun both subtend an angle of about 30 arc minutes, as seen from the Earth.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

development of X-ray astronomy by the close cooperation of X-ray, optical, and radio astronomers.

X-Ray Astronomy in 1970

Let me review a few of the highlights in the rapid development of X-ray astronomy that took place before satellites became available to X-ray astronomers in late 1970.

Most X-ray observations in the range of photon energies from about 1 to about 20 kev were carried out by means of sounding rockets; however, the early free-spinning rockets were soon replaced by rockets whose instruments could be pointed in a specified direction, or made to sweep slowly across a specified band of the sky. For the observations of X-rays of higher energies, which can penetrate the uppermost layers of the atmosphere, George Clark, in 1964, had introduced the use of high-altitude balloons.

The detectors used in the rocket experiments were still thin-window counters, except that the gas filling and the voltage were chosen so that the counters operated as *proportional counters*; among other advantages, these counters produced electric pulses of a size approximately proportional to the energy of the X-ray photons. Also, the area of the thin windows had been considerably increased.

The detectors used in the balloon flights were mainly *scintillation detectors*, which are particularly useful for the detection of photons with energies greater than about 15 kev. They produce light pulses which are again approximately proportional to the energy of the photons impinging upon them. Photomultipliers convert the light pulses into electric pulses. Measurements of the distribution in size of pulses from the proportional counters and scintillation counters made it possible to determine, if not the fine details, at least the general shape of the X-ray spectrum of various sources.

Great progress was achieved also in the design of the collimators, instruments used to define the field of view. An ingenious device (the *modulation collimator*) was invented by Minoru Oda in 1965 while working at MIT. To describe its operation would take us too far afield. Suffice to say that it combined a wide field of view with a fine angular resolution, which means that its characteristics were akin to those of a wide-angle photographic lens capable of producing pictures showing fine details of the subject. Widely used to this day, modulation collimators were an essential component of the instrumentation of the rocket which, in 1966, led to the identification of Sco X–1.

The exclusive use of rockets and balloons had imposed severe limitations: rockets provide only a few minutes of observing time and balloon-borne detectors can be used only to observe X-rays of fairly high energies, whose intensity, for most X-ray sources, is rather small. Despite these drawbacks, important results had been achieved. Extensive rocket surveys carried out mainly by the X-ray teams at NRL, at AS&E, at MIT, at the Lockheed Corporation (headed by Philip Fisher), at the Lawrence Livermore Laboratory (headed by Frederick Seward), and at Leicester University (headed by Kenneth Pounds) discovered about 40 X-ray sources. Several sources were discovered also by means of balloons. Most of these X-ray sources clearly belong to our

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

own galaxy, as shown by their clustering around the galactic equator. A beginning was also made in extragalactic X-ray astronomy. In 1966 observations by the NRL group, confirmed by later observations, established the existence of a weak X-ray flux from the Virgo Cluster, containing the radiogalaxy M–87. The NRL group also discovered that the radiogalaxy Cen–A, located in the constellation Centaur, and the Seyfert galaxy NGC1275 (in the Perseus Cluster) were X-ray sources.

Among the galactic sources, half a dozen (including the Crab Nebula) were identified with supernova remnants, having angular dimensions ranging from a few minutes of arc to a few degrees; the others appeared point-like, within the angular resolution of the detectors, and were generally designated as X-ray stars. In 1970 the only X-ray star definitely identified with a visible object was still Sco X–1. Evidence for a variability of some X-ray stars had been found in discrepancies between measurements in separate rocket or balloon flights, as noted first by Philip Fisher, and proof of rapid variability had been obtained in 1967 by Walter Lewin and George Clark in a balloon observation of a 20-minute flare of Sco X–1.

Moreover, two so-called transient X-ray sources had been discovered in the constellation of Centaur (Cen X–2 and Cen X–4); the first in a joint observational program of the University of Adelaide and the University of Tasmania, the second by a group at the Los Alamos Laboratories. Both of them appeared suddenly in regions of the sky where no X-ray sources had been previously detected, increased rapidly in intensity, becoming, in a week or so, as bright as the brightest steady X-ray sources, and then gradually faded away.

At the center of the Crab Nebula, a *pulsar* (pulsating radio source; more about these objects later) had been discovered in 1968. It has the highest pulsation frequency of all pulsars known to date (about 30 pulses per second) and it is the only one in which visible pulsations, synchronous with the radio pulsations, have been observed. Shortly after the discovery (early in 1969) of the visible pulsations, two independent rocket flights, the first by the NRL group, the second by a MIT group led by Hale Bradt, discovered pulsations in the X-ray part of the spectrum, from about 1 to about 10 kev. Some weeks later, a new analysis of balloon data obtained in 1967 by the X-ray group at Rice University and a balloon flight performed by a MIT group had extended the range of the X-ray pulsations to still greater energies. More recent experiments by the Cornell group have shown strong pulsations up to gamma-ray energies of 700 mev, while other experiments by Giovanni Fazio and Jonathan Grindlay at the Smithsonian Astrophysical Observatory, utilizing air showers produced by very high energy gamma rays, have provided evidence of pulsations about 10^{11} ev.

Finally the existence of a general diffuse background radiation (which had been already indicated by the 1962 rocket observations) had been firmly established. Above a few kev this radiation was found to be essentially isotropic, that is, having the same intensity in all directions, a strong argument in favor of its extragalactic origin. (A galactic origin would manifest itself by a maximum of intensity along the galactic equator).

On the other hand, evidence had begun to emerge for a galactic origin of most of the lower energy diffuse radiation. The most convincing confirmation of

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

this conclusion was produced a few years later by a rocket flight conducted by a group at the University of Wisconsin under the leadership of William Kraushaar. The experiment failed to detect the "shadow" of the Small Magellanic Cloud which would have been expected if the low-energy X-ray background radiation originated beyond this object.

So much for the observational picture. What about the interpretation, particularly in regard to the galactic sources for which a fair amount of observational information was already available?

The supernova remnants and the X-ray stars posed two entirely different problems. In the first instance, the nature of the object was, in a general way at least, well known, and the question was only to determine the mechanism responsible for the X-ray emission. In the second instance, however, we were confronted with the existence of a previously unknown class of celestial objects, whose outstanding characteristic was that a large proportion of the radiated energy was in the form of X-rays.

In order to fully appreciate the physical significance of the findings of X-ray astronomy (both those already presented and those to be described in the remainder of this paper) it will be necessary for the reader to recall some of the notions and ideas stemming from previous experimental and theoretical work in physics and astrophysics.

An ordinary star, such as the Sun, is heated by the "burning" of nuclear fuel; the *fusion* of hydrogen to form helium, the subsequent fusion of helium and of the elements thus formed to produce heavier elements. The nuclear energy is released in the interior of the star and travels gradually toward the surface. The heat thus generated produces a pressure which counterbalances the gravitational attraction.

When the nuclear fuel is nearly exhausted, the star, not being completely supported any longer by the thermal pressure, begins to collapse, releasing gravitational energy in the form of heat and perhaps shedding some of its mass. At a certain stage, however, a new effect, called *Fermi pressure*, comes into play, which tends to oppose further collapse. This phenomenon, unknown to classical physics, is rooted in the basic principles of quantum mechanics. For those of my readers who are not familiar with this branch of science, I shall state only that certain kinds of particles (electrons, protons, and neutrons among them) refuse to be packed beyond a certain density unless they are supplied with a certain amount of energy, which increases as the density increases, and which is greatest for the lightest particles, i.e., for the electrons. In the relatively diluted matter of ordinary stars, the Fermi pressure of the electrons is negligible compared with the thermal pressure. As the star contracts, however, the Fermi pressure will increase rapidly and, if the residual mass does not exceed about 1.4 solar masses, it will eventually stop the collapse altogether. The end product will be a *white dwarf*.

White dwarfs have been known to astronomers for about a century (although the cause of their stable equilibrium, the Fermi pressure of electrons, was not understood until the middle 1920s). Their average dimensions are of the order of those of the Earth, about 100 times smaller than those of the Sun.

However, when the residual mass is greater than about 1.4 solar masses, the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

collapse does not stop at the stage of a white dwarf. The reason is that electrons are not indestructible particles. Under enormous pressures, they may combine with protons to form neutrons. In this case, then, the collapse proceeds until a new stable configuration, known as a *neutron star*, is reached. In such an object the bulk of the matter consists of free neutrons, and it is the Fermi pressure of the neutrons, together with the nuclear-force repulsion between neutrons at close distance, which counterbalances the gravitational forces. The typical radius of a neutron star is believed to be of the order of 10 kilometers.

Neutron stars, predicted by the theorists in the thirties, became an observational reality in 1968 with the discovery of the pulsars, and with the overwhelming evidence, on which I cannot dwell here, that these are indeed magnetized rotating neutron stars producing pulses by a mechanism similar to that of a rotating lantern in a lighthouse.

According to Einstein's theory of gravity, there should exist an even more condensed state of a collapsed star. In fact, when the residual mass exceeds some critical value, the collapse cannot stop at the stage of a neutron star, but must continue until eventually the star disappears behind a surface where the gravitational potential is so large that nothing can escape from within it, not even photons. This critical surface is known as a *black hole*, because it swallows everything that falls upon it, without possibility of reemission.

The critical mass is estimated theoretically to be about 2 solar masses. For such a mass, the radius of a black hole is about six kilometers; this radius varies in proportion to the mass.

Black holes appear to be a theoretical necessity. But, as we shall see, observational evidence for their existence was obtained only recently from X-ray astronomy.

It is believed that many supernova explosions are the result of the collapse of an ordinary star into a neutron star (although some supernova explosions may produce black holes, and others may result from nuclear detonation, in which case the star would blow apart completely). Pulsars, that is, rotating neutron stars, have been detected in the Crab Nebula and in Vela X, a supernova remnant about ten times older than the Crab Nebula. The gravitational energy released in the collapse into a neutron star is truly enormous. It amounts to ten times the maximum energy that could be generated by the nuclear burning of the corresponding mass (assumed to consist originally of pure hydrogen). Part of the energy is changed into kinetic energy of the matter ejected during the collapse. Thus a fast-expanding cloud is produced. What energy is left over remains stored in the neutron star, as kinetic energy of rotation (from the basic principle of the conservation of angular momentum, it is known that if, before collapse, the star was rotating even at a slow rate like our Sun, the rotational velocity will reach a very large value once the collapse is completed).

Some discussion of radiation processes may also be useful at this point.

Solar matter, like that of other ordinary stars, is very opaque to electromagnetic radiation of all wavelengths. Therefore the bulk of solar radiation comes from a thin surface layer (the photosphere) and its spectrum closely resembles that of a *black body*, a surface that absorbs all radiation falling on it with no reflection. For a temperature of about 6000° K, such as that of the solar surface, the spectrum peaks in the visible and falls virtually to zero in the far

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

infrared and in the far ultraviolet. In the solar atmosphere above the photosphere the temperature first drops slightly, then rises rapidly through the narrow layer known as the chromosphere until it reaches, in the extended and tenuous solar corona, values of the order of millions of degrees. Solar X-rays originate primarily in the lower corona; but even there the density is so low that only a comparatively small flux of X-rays is produced (by thermal bremsstrahlung; see below). A similar situation prevails in most ordinary stars.

In addition to black body radiation, possible emission processes to be taken into account in the discussion of celestial X-ray sources include *bremsstrahlung*, the *synchrotron effect*, and the *inverse Compton effect*. The first two effects are a direct consequence of the classical theory of electromagnetism, which asserts that electromagnetic waves are radiated whenever a moving electron changes either the direction or the magnitude of its velocity (depending on the particular conditions, the electromagnetic waves will belong to different parts of the spectrum, such as the radio, the visible, and the X-ray ranges; the waves emitted by a radio station are caused by electrons moving back and forth along the transmitting antenna). In the case of bremsstrahlung, the radiation arises from the deflection of fast electrons during close encounters with atomic nuclei; in the case of the synchrotron effect, the radiation arises from the deflection of fast electrons crossing a magnetic field. The inverse Compton effect is not a classical, but a quantum phenomenon, and one can best describe it by considering a radiation as a stream of photons. One may then think of the inverse Compton effect as a collision of a low-energy photon with a high-energy electron, resulting in a loss of energy by the electron and a corresponding gain of energy by the photon. In this manner, an infrared photon, for example, may be raised into the X-ray spectral range.

In all these processes, of course, the spectrum of the emitted radiation depends on the energy distribution of the electrons. Particularly important, it turns out, is the case of *thermal bremsstrahlung*, that is, bremsstrahlung by electrons moving about at random in a hot plasma that is sufficiently thin to be transparent to its own radiation. One can easily show that, if the plasma is in thermal equilibrium at a constant temperature, the spectrum of its thermal bremsstrahlung has a particular shape, dropping rapidly toward zero beyond a photon energy equal to the kinetic energy of the plasma particles. The bulk of the radiation will be in the form of X-rays if the temperature exceeds several million degrees.

Considering first supernova remnants, the prevailing view in 1970, which has not changed since, was that the expanding cloud generated in the initial explosion, interacting with interstellar matter, gives rise to a "shock," which results in a shell of very hot gas. In almost all supernova remnants, this is the source of the visible and of the radio emission, which, in fact, exhibits the shell structure required by this interpretation. The unexpected discovery of a strong X-ray emission meant that the density of the gas (at least in some portion of the shell) was higher than had been previously assumed. This is an important piece of information that substantially enlarges the observational basis for the quantitative interpretation of the shock formation, a very difficult problem in the theory of very dilute gases. (A joint rocket experiment conducted later by the X-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

ray groups of AS&E and of Columbia University showed that the X-ray emission from the supernova remnant known as the Cygnus Loop has also a shell structure similar to that of the optical and radio sources).

An exception is the Crab Nebula. This is a strong radio source. Its optical and radio emission does not come from a shell, but from the whole cloud. There is convincing evidence that this radiation is due to the synchrotron effect by high energy electrons circulating in the magnetic field that permeates the cloud.

It was natural to assume that the X-ray emission (whose spectrum appeared to be a continuation of the radio and visible spectra) was caused by the same effect. This assumption was confirmed later by rocket experiments performed by the Columbia group under the leadership of Robert Novick, which showed that X-rays, like optical and radio waves, are strongly polarized, a property characteristic of synchrotron emission.

The discovery that the synchrotron emission of the Crab Nebula extends into the X-ray range had an important consequence. The electron energy needed to produce X-ray photons is considerably greater than that needed to produce the much less energetic photons of ordinary light (over one hundred thousand billion ev, as against about one thousand billion ev). The theory of the synchrotron effect shows that the greater the electron energy, the faster is the electron energy loss. Whereas the electrons responsible for the optical and radio emission have an effective mean life of the order of one thousand years or more, and thus could have been produced at the time of the initial explosion, those responsible for the X-ray emission have a mean life of only several years and even less. Therefore the Crab Nebula must contain an "accelerator" capable of continuously producing electrons of exceedingly high energies (enormously greater than the energies reached by existing or planned terrestrial accelerators). The nature of this accelerator is not yet known in detail. However, there is little doubt that the source of the required energy resides in the rotating neutron star left behind by the supernova explosion. The rotational period of this star had been found to increase at the rate of one part in 2400 per year. The corresponding loss of rotational energy closely matches the total energy radiated by the Crab Nebula (with something left over for the possible acceleration of cosmic-ray particles).

As for the pulsating X-rays emitted directly by the Crab pulsar, the details of the production process were (and still are) not yet firmly established (as they are not known for the pulsating visible and radio waves); except that this process is certainly not of a thermal nature, but is rather akin to a synchotron process.

Let us next turn to X-ray stars. As I have said, their nature in 1970 was completely unknown. Many of these stars, and in particular Sco X–1, exhibit X-ray spectra similar to those predicted for a thermal bremmstrahlung. Therefore they were thought to contain hot plasma clouds. A detailed analysis of the X-ray, optical, and infrared data of Sco X–1, for example, suggested a plasma cloud of about 10^8 cm radius ($1/6$ of the Earth's radius), a density of about 10^{16} electrons per cm^3 (air at sea level contains about 10^{19} atoms per cm^3), and a temperature of several tens of millions of degrees. But how were such clouds formed, how were they kept together, or continuously replenished, and how were they supplied with the necessary energy?

Hypotheses, of course, were not lacking. One rather widely held view was

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

that X-ray stars were binary or more complex systems (i.e., systems composed by two or more stars), with a white dwarf or a neutron star (i.e., a collapsed object) as one of the partners. Matter would accrete from the ordinary star (or stars) to the collapsed object, which, having a mass similar to that of the Sun, but much smaller dimensions, is surrounded by a gravitational field much stronger than that surrounding ordinary stars; that is to say, it lies at the bottom of a very deep gravitational "potential well." Matter falling into this well would release a large amount of gravitational energy, which, in some way or other, would be transformed into heat. However, no clear experimental evidence supporting this model was available.

UHURU

A new chapter in the history of X-ray astronomy began in December 1970, with the launching from an Italian platform off the coast of Kenya of SAS-1, the first in the NASA series of Small Astronomical Satellites, and the first satellite specifically designed for and entirely devoted to X-ray astronomy. The launching took place on Kenya's independence day, and, because of this coincidence, the satellite was named UHURU, the Swahili word for "freedom." The instrumentation for this satellite was built at AS&E under a NASA contract, by Giacconi's group.

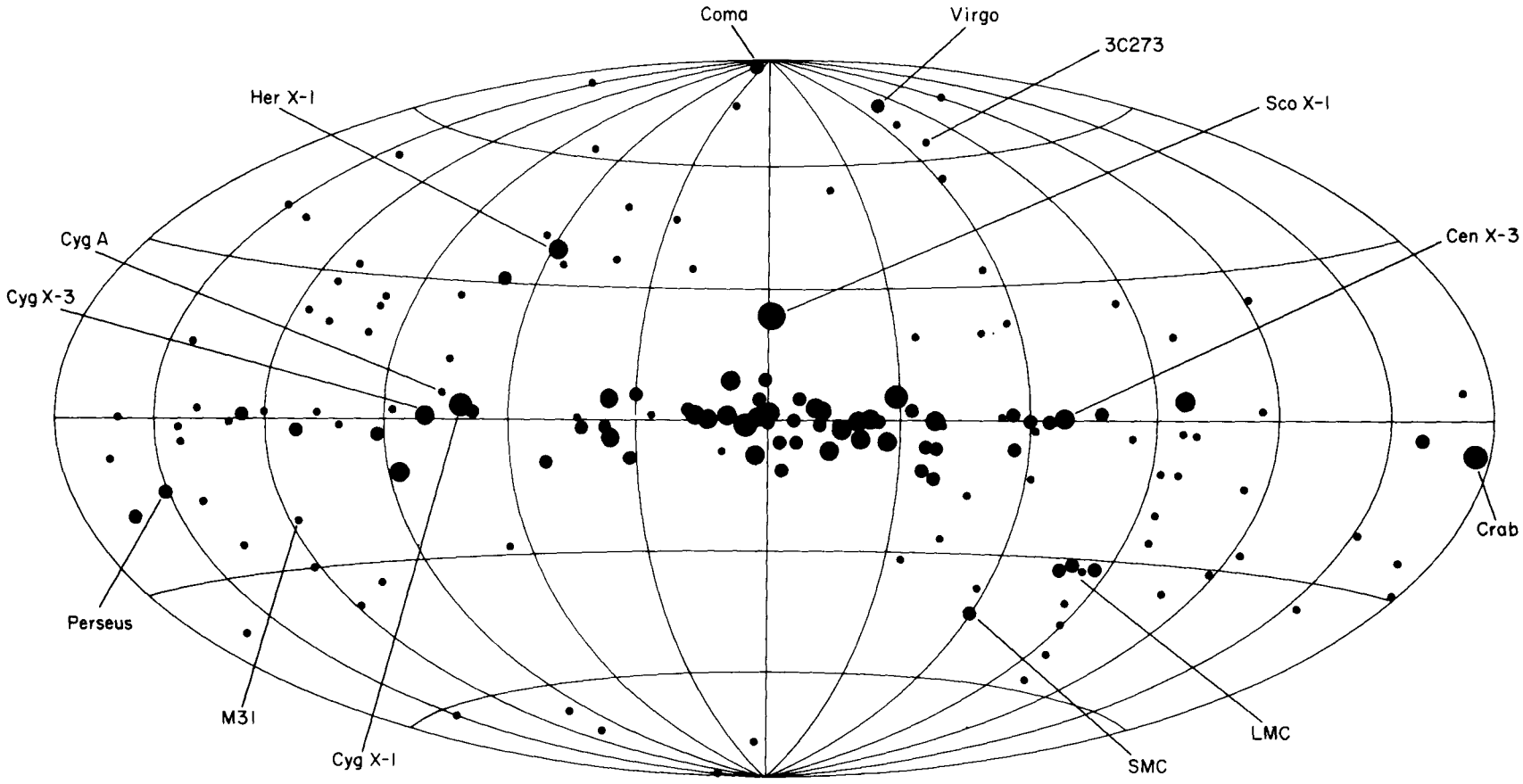
I shall not attempt to describe here the scientific payload and the operation of UHURU, except for pointing out two essential features. The spin axis of this satellite could be oriented at will by command from the Earth. The spin rate could also be controlled, from several revolutions to less than one revolution per hour; so that the X-ray detectors (proportional counters with large window area), looking perpendicularly to the spin axis, would sweep slowly and repeatedly across a chosen band of the sky. The purpose of these features was to enable UHURU to accomplish what was conceived as its main assignment: to discover new, weak X-ray sources and to determine the positions of these and of the already known sources with sufficient accuracy to facilitate the identifications of X-ray sources with visible objects.

This assignment was fulfilled with remarkable success. The third UHURU catalogue contains 161 sources; their celestial distribution, in galactic coordinates, is shown in Figure 1. Of these 161 sources, about 100, which include the brightest objects, are clustered around the galactic equator, particularly toward the galactic center, and belong undoubtedly, for the most part, to our galaxy. The remaining, weaker sources show a more or less uniform directional distribution, such as expected for extragalactic objects.

As a very important result of this survey, UHURU confirmed the few extragalactic X-ray sources discovered by previous rocket flights, and detected X-ray emission from several additional extragalactic objects. The list of these objects was thus expanded to include a spiral galaxy similar to our own, two radio galaxies (galaxies with large lobes of radio emission), one isolated Seyfert galaxy (a galaxy which, from optical observations, appears to have a particularly active core), several clusters containing from a few hundreds to a few thousands of individual galaxies, and one quasar (3C273). Moreover, UHURU's observations showed that the X-ray flux from the radio galaxy Cen-A does not originate

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED



A map of the X-ray sky in galactic coordinates derived from the 3U Catalog. The location of each X-ray source is approximately shown. The size of the dots is proportionate to the logarithm of the intensity. *Astrophysics Journal* Supplement Series No. 237, 27:(1974) 37–64.

from the extended radio lobes (as it had been suggested) but from a small region containing the active core of the visible galaxy. Finally, UHURU proved, in at least three cases, that clusters are extended X-ray sources.

This is an impressive list of accomplishments.

Yet perhaps the most striking result of UHURU's observations was one that goes beyond its original assignment. I refer here to the unambiguous proof that several X-ray stars are indeed binary systems, as had been tentatively suggested previously.

Early in its life, UHURU discovered that two X-ray stars, Cen X–3 and Her (Hercules) X–1, were regularly pulsating, a discovery made possible by UHURU's capability of scanning a given source slowly and repeatedly.‡ The pulsation periods were about 4.84 sec and 1.24 sec, respectively, and could be measured with very high accuracy. The binary character of Cen X–3 and Her X–1 was established by the observation of periodic eclipses, coupled with the observation of Doppler shifts of the pulsation periods. (Because of the orbital motion of the X-ray source, the apparent pulsation period shifts back and forth, as the source moves toward or away from us. A similar effect produces periodic apparent changes in the wavelength of the spectral lines of a visible star in a binary system.§) The orbital periods were found to be very short for both Cen X–3 (2.07 days) and Her X–1 (1.7 days), indicating close binary systems. The companion stars were soon discovered by optical observations. The companion

‡For the history, I note that Walter Lewin and his associates at MIT, in a balloon flight carried out in October 1970, had discovered a new source (GX1+4) which appeared to pulsate with a period of about 138 seconds. Five years later when the source was seen again by the satellites Ariel–5 and SAS–3, its pulsation period had considerably shortened, to about 120 sec.

§The Doppler effect (so named after the Austrian scientists who discovered it in 1842) is of basic importance in all fields of observational astronomy. The following explanation is added here for the benefit of those among my readers who are not already familiar with it and wish to acquire a clear understanding of its nature.

Think of a clock mechanism, some distance away, which sends a light flash exactly once every second. Because of the finite velocity of light (close to 300,000 kilometers per second) we will see the flashes with some delay; however, if the clock is at rest, the flashes will arrive still exactly one second apart. But suppose now that the clock is moving away from us at a speed of, say, 300 kilometers per second (not unreasonable if the "clock" is, in fact, a celestial object). Then, between two flashes, its distance from us will have increased by 300 km; the second flash will have to travel 300 km more than the first, and will arrive with a further delay of one millisecond. Therefore the time between flashes will appear to us to be 1.001 sec rather than just one second. Similarly, if the clock is moving toward us with the speed of 300 km per second, the time between flashes will appear to be 1-0.001 = 0.999 sec. For the same reason, the spectral lines of a light source will appear displaced toward the red (longer period, longer wavelength) if the source is moving away from us, or toward the blue (shorter period, shorter wavelength) if the source is moving toward us. (A familiar consequence of the Doppler effect for sound waves is the sudden decrease in pitch of the horn of a fast-moving car, when the car overtakes and passes us).

Think now of a system consisting of an ordinary star and a pulsating collapsed object revolving around their common center of mass. Unless the orbits of the stars lie in a plane perpendicular to the line of sight, the pulsation period of the collapsed object will appear slightly shorter than their true period while the object is approaching, slightly longer while receding. If moreover the so-called inclination angle, i.e., the angle formed by the orbital plane with the plane of the sky (which is the plane perpendicular to the line of sight) is not too small, the collapsed object will periodically disappear behind the ordinary star; i.e., will undergo periodic eclipses. The spectral lines of the ordinary star will also show periodic displacements toward the red and toward the blue. Because the velocities of the collapsed object and of the normal star around their common center of mass are inversely proportional to their masses, a comparison between the amplitudes of the Doppler effects of the spectral lines and the X-ray pulsations will yield the ratio of the masses of the two stars.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

of Cen X–3 turned out to be a "blue supergiant," that is, a very active, massive, young star; that of Her X–1 appeared to be a low-mass star, somewhat similar to our Sun.

The collapsed nature of the X-ray sources is experimentally demonstrated by the short pulsation periods, which rule out ordinary stars. On the other hand, for a number of reasons (some having to do with the disrupting effects of centrifugal forces) it appears very unlikely that the collapsed objects are rotating white dwarfs. Nor does it seem probable that white dwarfs can undergo stable vibrations with the observed short periods. Black holes are not regarded as capable of producing regular pulses. Therefore neutron stars, such as those previously recognized as being responsible for the pulsars, appear to be by far the most likely choice. It is believed that these neutron stars are strongly magnetized. The magnetic field lines guide the accreting material toward the magnetic poles, so that the X-ray emission is strongest from the polar regions. Rotation of the neutron stars around an axis different from the magnetic axis produces the observed pulsations.

The discovery of X-ray sources in binary systems, first published in early 1972 by the UHURU group, was of fundamental importance on two accounts. (1) It provided an explanation for the operation mechanism of at least some X-ray stars, and (2) it made it possible to estimate the masses of the collapsed objects.

A word of caution is necessary at this point. To achieve even a moderate degree of precision in the determination of the mass of the collapsed partner in a binary system is a very hard task. One of the major difficulties lies in the measurement of the Doppler effect of the spectral lines of the visible partner, which is needed to supplement the information derived from the accurately measurable Doppler effect of the X-ray pulsations.

Of potentially still greater consequence than the discovery of the binary nature of Cen X–3 and Her X–1 was recognition of the binary nature of another X-ray star, Cyg (Cygnus) X–1. This strong source had been known for several years, but all attempts at its visual identification had failed (in part because Cyg X–1 lies in a densely populated region of the sky). When striving to reduce the area of uncertainty for the position of Cyg X–1, UHURU, quite unexpectedly, detected the occurrence of very fast fluctuations in its X-ray emission. Shortly thereafter a MIT rocket observation showed that the fluctuations occurred at random, and further improved the determination of the position made by UHURU. Following these measurements, radio signals originating from Cyg X–1 were detected by radio astronomers at the Radio Observatory in Westerbork in The Netherlands and at the National Radio Astronomy Observatory in the United States. The radio position was accurate to a few arc seconds, and as a consequence Cyg X–1 could be reliably identified with a "blue supergiant." The discovery that Cyg X–1 was a binary system was the result of careful observations of the visible star (by B. Webster and P. Murdin and by C. Bolton), which showed that its spectral lines underwent the periodic Doppler shifts characteristic of a star in a binary system. The orbital period was found to be 5.6 days. Note that this result could not have been obtained from X-ray observations because the X-ray source did not undergo eclipses and the rapid variations of its X-ray intensity, not being in the form of regular pulsations, did

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

not lend themselves to the detection of a Doppler effect. The absence of X-ray eclipses simply meant that the line of sight was not sufficiently close to the orbital plane.

Great efforts were made to estimate the mass of the collapsed object on the basis of the optical data. The results are still very uncertain, but they do place for this mass a lower limit of about five solar masses, which exceeds the theoretical value for the maximum mass of a neutron star. One is thus led to the conclusion that the collapsed object in Cyg X–1 is a black hole. As I already mentioned, black holes, until that time, had been merely a theoretical conjecture, and the experimental demonstration of their existence would be a result of major importance. It is thus natural that the evidence derived from the study of Cyg X–1 should be subjected to a most critical analysis. I cannot go into details here. I can only express my personal view, which, I believe, is shared by the majority of my colleagues; namely, that this evidence is fairly strong, but does not yet amount to a certainty. Therefore further investigations, and particularly the search for other objects similar to Cyg X–1, is most desirable.

The Post-URUHU Era

After UHURU, a number of other satellites were launched. Two of them were devoted entirely to X-ray astronomy (the British satellite Ariel–5, in 1974, and the American satellite SAS–3, in 1975). Others were designed to carry out X-ray observations as part of a wider astronomical program.

Ariel–5 is a cooperative project of the X-ray groups at the University of Leicester (headed by Kenneth Pounds), the University of Birmingham (headed by Peter Willmore), the University College, London (headed by R. L. Boyd), and the Imperial College, London (headed by H. Elliot). In addition it also carried an all-sky monitor for strong transients provided by Steven Holt of the Goddard Space Flight Center.

The instrumentation of SAS–3 was built at MIT by the X-ray group headed by George Clark.

Most of these satellites are still active and carry payloads whose capability exceeds that of the relatively simple instrumentation of UHURU. Most importantly, some (SAS–3 among them) are provided, unlike UHURU, with three-axis stabilization, so that their instruments can be pointed steadily, for long periods of time, to a chosen target.

The number and variety of data provided by the new satellites, as well as by some recent sounding rocket and balloon flights, are such that any attempt to summarize them all would be an impossible task. So I shall restrict my discussion to a small sample of the highlights.

Eclipsing Pulsating Binaries

The list of the known eclipsing pulsating binaries has been increased from two to four with the addition of SMC X–1 (in the Small Magellanic Cloud) and of Vela X–1. Critical analysis of all X-ray and optical data for the four binaries has led to the results shown in Table 1.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Table 1

Orbital periods, pulsation periods, and masses (in solar masses) of the collapsed objects in the four known eclipsing and pulsating X-ray binary systems (from a review paper by Paul Joss and Saul Rappaport).

Name	Orbital Period (days)	Pulsation Period (seconds)	Mass of Collapsed Object (solar masses)
Cen X-3	2.087	4.842	0.7 to 4.3
Her X-1	1.700	1.238	0.0 to 2.3
SMC X-1	3.892	0.715	0.8 to 1.8
Vela X-1	8.966	283	1.2 to 2.4

Note that the data available at present are compatible with a unique mass between 1.2 and 1.8 solar masses for all four neutron stars.

The allowed range of masses of neutron stars depends on the still unknown interactions among nucleons at very close distances. The study of X-ray binaries appears to offer, as I have already pointed out, the best available source of information about these masses. Hence the great importance for nuclear physics of improving the mass determinations in pulsating binaries. These measurements have also an important bearing on the problem of stellar evolution, and great interest is attached to the question of whether or not all neutron stars have approximately the same mass, irrespective of the mass of the progenitors.

Variability of X-Ray Stars

I have already noted that, since the early days of X-ray astronomy, some X-ray sources were found to undergo temporal variations in their X-ray emission. Many more examples of variability were discovered by UHURU (in addition, of course, to the regular pulsations and eclipses of X-ray binaries caused by the rotational and the orbital motions of the collapsed components).

However, it was not until the more refined observational capabilities of some of the new satellites became available that temporal variations came to the foreground of X-ray astronomy. They now appear to be a general property of most, if not all, galactic X-ray sources other than supernova remnants, making the X-ray sky dramatically different from the relatively quiet visible sky. They exhibit a bewildering variety with regard to time scale, magnitude, and structure. It has not yet been possible to classify them into sharply distinct categories; perhaps it will never be possible to do so, because it is likely that most or all of these variations have their ultimate origin in the exceedingly complex phenomena occurring in the plasma that surrounds an accreting collapsed object, phenomena which influence the rate of accretion, as well as the absorption and the scattering of the X-ray flux originating from this object.

One particular kind of variable X-ray stars are the *transient sources*, sometimes called X-ray novae, of which two had been discovered prior to the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

launching of UHURU, as discussed earlier. Several more were discovered by UHURU and by OSO–7, a satellite launched shortly after UHURU, whose main assignment was to study solar X-rays, but which carried also two X-rays detectors, one developed by the group at the University of California in San Diego headed by Peterson, the other by the MIT group headed by Clark. The list of transients was considerably increased by the more recent satellite observations. The X-ray detector of the Leicester group aboard Ariel–5 discovered in one year about a dozen new transients. Two were particularly interesting, because they exhibited regular pulsations in their X-ray emission. One of them, the Taurus–X Nova, rose to the intensity of twice the Crab Nebula in about 10 days and then became slowly dimmer, with occasional small brightenings, so that it could be observed for a period of several months.

Efforts were made, by SAS–3, to detect temporal changes of the pulsation period, that may be indicative of a periodic Doppler effect, and therefore of an orbital motion. The results are still of a preliminary nature. However, they suggest that the object may be a binary system, with a very long orbital period, of several months perhaps. Optical observations suggest again a supergiant as the visible partner. Thus it would seem that the Tau–X Nova consists of a rotating neutron star (or white dwarf?) orbiting a supergiant at a fairly large distance and accreting matter from its partner only on rare occasions.

Another remarkable transient was the Monoceros Nova; it is the only X-ray nova to be identified with an optical nova, that is with a visible star that increased in luminosity by a large factor at the time of the sudden appearance of the X-ray star. Some evidence was produced for a small modulation, in its X-ray emission with an 8-day period, presumably caused by an orbital motion, which would indicate that the Monoceros Nova, too, belongs to a binary system.

The most striking, entirely new result concerning the variability of X-ray stars has been the recent discovery of the *bursting sources*. These are sources whose characteristic property is the production of X-ray "bursts," which rise to a peak intensity occasionally greater than that of the Crab Nebula in a second or less, and last usually about half a minute.

Two such bursts coming from different sources were first detected by the Soviet satellite Cosmos 428 in 1971 and were reported in 1975 by a group headed by A. S. Melioranskii. However, bursting sources did not attract general attention until early 1976, when Jonathan Grindlay and John Heise, who were not aware of the Russian observations, reported two bursts about 8 hours apart, detected by the Astronomical Netherlands Satellite (ANS) in September 1975, which appeared to originate in the globular cluster NGC6624 (a nearly spherical cluster containing about 100,000 stars).

Immediately after this announcement, Clark and his collaborators, reexamining records obtained in May 1975 when NGC6624 was under observation by SAS–3, discovered 11 bursts which occurred fairly regularly spaced with a separation of about 4.3 hours. During a subsequent observation period in February 1976, lasting several days, the source was not bursting. Finally during four days in March 1976, SAS–3 detected 22 bursts, whose separation changed gradually from about 3.4 hours to about 2.2 hours; afterward, the burst activity ceased and did not resume during the following four days of observation. In all

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

the observations performed so far, the steady source in NGC6624 was found to be in a low-intensity state when bursts were occurring and in a high-intensity state when bursts were absent.

In the meantime, the existence of a number of additional burst sources had been established (about 17) and new sources are still being discovered as these lines are written (November 1976). They are strung along the galactic equator, with a concentration near the galactic center. Observations performed by several other satellites and a new analysis of old UHURU data contributed to our present knowledge concerning these sources. However, so far, the most systematic investigation of burst sources was performed by a SAS–3 group led by Walter Lewin.

Lewin's group discovered 10 of the 17 sources mentioned above. Three of them lie within one-half of one degree from the galactic center. Another source (known as the Rapid Burster), which is located at 6 degrees from the galactic center, produces sometimes bursts at irregular time intervals ranging from a few seconds to a few minutes; the total flux in a burst is approximately proportional to the time interval to the next burst. This temporal behavior is drastically different from that of all other known sources, whose bursts are usually separated by time intervals of the order of hours or days.

For several sources periods of burst activity are known to alternate with periods of quiescence. No sufficient information is yet available to establish whether or not this behavior is common to all burst sources. Some burst sources, possibly all of them, are also steady X-ray emitters.

In the area of uncertainty determined by SAS–3 for the Rapid Burster, the optical astronomer, William Liller, detected a previously unobserved globular cluster. Therefore almost certainly two burst sources (the one in NGC6624 and the Rapid Burster) are located in globular clusters. The same may be true for other burst sources. However, it appears that most burst sources do not lie in globular clusters.

The results of SAS–3 strongly suggest that the origin of the bursts is some sort of relaxation phenomenon. Most likely the burst sources are collapsed objects (neutron stars or perhaps black holes) immersed in fairly dense plasma clouds. For some reason, not all of the matter sucked in by the gravitational field flows steadily to the collapsed object. Perhaps some fraction accumulates gradually at a fairly large distance. Eventually an instability develops, which breaks the "barrier" and allows the plasma to fall suddenly toward the collapsed object, releasing its gravitational energy in the form of an X-ray burst.

Extragalactic Sources

Post-UHURU satellites and rockets have also produced significant new data on extragalactic X-ray sources.

UHURU had discovered a number of very weak sources which, because of their large distance from the galactic equator, were good candidates for extragalactic X-ray emitters. However, insufficient accuracy in their position had prevented any identification. Many of these sources have now been identified by Ariel–5 with distant galaxies.

The association of X-ray emitters with active compact galaxies, established

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

by UHURU for the radio galaxy Cen–A and the Seyfert galaxy NGC4151, has been confirmed by the discovery of an X-ray emission by the Seyfert galaxy 3C120 by Herbert Schnopper and his associates of the Smithsonian Astrophysical Observatory working with SAS–3, and of a dozen more Seyferts by Pounds and his associates working with Ariel–5.

The question whether the extended X-ray sources in a galactic cluster are due to the emission from diffuse clouds or from many point-like objects has not yet been entirely resolved. However, two results bearing on this important question deserve special mention. The first is the identification by Ariel–5 of the spectral lines of highly excited iron ions in the Perseus Cluster (Leonard Culhane and his associates); a clear indication of the presence in this cluster of a plasma at a nominal temperature of the order of ten million degrees. The second is the demonstration that the Perseus Cluster contains both a diffuse source and a point-like X-ray source coincident with the active galaxy NGC1275. To this result have contributed a rocket experiment by Paul Gorenstein, observations with the satellites Copernicus¶ and SAS–3, and a rocket experiment by Richard Wolf of the Columbia University group.

Conclusions

Galactic X-ray astronomy has reached, by now, a degree of maturity where a great variety of manifestations begins to appear caused by accidental circumstances concealing, perhaps, a basic unity of causes.

Let us focus our attention on the strong point-like sources, which constitute the majority of the galactic sources, and which we may still want to call X-ray stars, although we know now that they include such seemingly diverse objects as pulsating and nonpulsating binaries, transients, bursters, etc. We are leaving out supernova remnants as well as several recently discovered weak X-ray sources coincident with comparatively nearby stars, well known to optical astronomers before the beginning of X-ray astronomy.

For a number of X-ray stars we know that the X-ray energy derives from the gravitational energy that is released when gas from the surrounding space falls into the deep potential well at the bottom of which lies a collapsed object. The best known examples are Cen X–3, Her X–1, SMC X–1, Vela X–1, and Cyg X–1. There is convincing evidence that each of these X-ray stars is a close binary system, consisting of an "ordinary" star and a collapsed object (most likely a neutron star in four of them, probably a black hole in the fifth); and that matter accretes from the atmosphere of the ordinary star onto the collapsed object.

There exist several other X-ray stars whose binary nature is strongly suggested by periodic intensity variations but for which it has not been possible to construct a simple and clear model. Sco X–1 is one of them. Recent optical observations have revealed a light variation with a regular period of 0.787 day, suggesting an orbital motion. However, no similar X-ray periodicity has been observed as yet. Curiously enough, Sco X–1, the brightest permanent X-ray

¶ A satellite designed primarily for extreme ultraviolet observations, but equipped also with grazing incidence concentrators, the latter a joint project of the University of Leicester and the University College of London.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

source in the sky, the one whose discovery in 1962 opened up the new field of X-ray astronomy, and the one, perhaps, which has been studied in greatest detail (especially in the early years), is still one of the least understood of all X-ray sources.

Among the transient sources, one at least (the Tau–X Nova) may be, as we have seen, a binary system. The transient behavior would be accounted for by sporadic accretion phenomena. It is likely that a similar interpretation applies to the other transients.

Bursting sources, while initially appeared to be objects of an entirely different kind from the other X-ray stars, now are generally believed to be also powered by accretion except that the particular conditions prevailing in the neighborhood of these objects cause the flow to occur intermittently, in a manner suggesting relaxation oscillations. Whether the accreting matter is the atmosphere of an ordinary star, forming a binary system with the collapsed object, or whether it is a dense cloud surrounding this object, is still under-debate. However, despite the fact that, as I already noted, no observational evidence for the binary nature of bursting sources has yet emerged, the balance seems somewhat tipped toward the first assumption.

Are we then to conclude that all X-ray stars (which do not include supernova remnants or weak X-ray sources in previously known visible objects) derive their energy from accretion onto collapsed objects? I suppose that most X-ray astronomers would tentatively answer in the affirmative. For one thing, it would be difficult to find another equally efficient source of energy. It is appropriate to stress here the truly enormous power output of X-ray stars. It ranges from about one hundred to over a million times the total power output of the Sun.

However, the conclusion that gravitational energy of accretion is the only source of energy for X-ray stars is far from a certainty. Only a minority of X-ray stars have been studied so far in detail and further, more extended, observational data are most desirable to decide whether or not processes other than accretion are at work in some X-ray stars. Moreover, what I have called "accidental circumstances" are by no means a trivial matter. They involve the most difficult and fundamental problems concerning the theory of very dilute plasmas, where direct collisions between particles play a negligible role and the collective behavior is determined by the long-range electric and magnetic forces. In principle at least, X-ray astronomy offers a possibility to investigate these phenomena under conditions far beyond those that can be achieved on the Earth.

Finally, as we have seen, further, more accurate X-ray and optical studies of X-ray binaries are of basic importance as the most promising method for improving our knowledge of the masses of neutron stars and possibly for establishing the existence of black holes.

Extragalactic X-ray astronomy is in a much less advanced state of development than is galactic X-ray astronomy.

To proceed further, more sophisticated instruments than those used so far will be needed, and here is where we expect that grazing incidence optics will come into the foreground. An X-ray telescope of sufficient area and provided with sufficiently sensitive detectors to produce images of extragalactic X-ray

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

sources with reasonable exposure times has been constructed by AS&E under the direction of Giacconi and his coworkers, now at the Center for Astrophysics of the Smithsonian Institution and Harvard University. It will have an angular resolution of the order of seconds of arc, similar to that of optical telescopes. It will thus be capable of distinguishing fine details of the X-ray sources in active galactic nuclei and globular clusters, and to relate the structure of these sources to those of the optical and radio sources. It will yield results relevant to the problem concerning the fraction of the general isotropic X-ray background due to very faint, hitherto undetected sources.

Auxiliary instruments (including a low resolution objective grating spectrometer and a high resolution crystal spectrometer) have been provided by the Center for Astrophysics, by the Center for Space Research of MIT, and by the Goddard Space Flight Center of NASA.

The satellite assigned by NASA to the X-ray telescope, High Energy Astronomy Observatory B (or HEAO–B), is scheduled for flight in 1978. It will be the second of the HEAO series; the first, HEAO–A, a scanning satellite provided with various systems of counters and collimators of very large sensitive area, up to about one square meter, will be launched in 1977.

The operation of HEAO–B will be under the control of a consortium consisting of the three centers mentioned above and of the Astrophysics Laboratory of Columbia University. Guest investigators will be invited to participate in the observational program. There is every reason to expect that the launching of the HEAO satellites will signal a new qualitative leap forward in the development of X-ray astronomy.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

SHMUEL N. EISENSTADT

Sociological Theory and an Analysis of the Dynamics of Civilizations and of Revolutions

The Dynamics of Societies and Civilizations in Classical Sociology and in Contemporary Studies of Modernization

I

SOCIOLOGY IS IN A PERIOD of great turmoil, in which basic premises and approaches are being reexamined and attempts at consolidation are being made. This has in turn reopened the fundamental problems of sociological analysis.[1] In this paper, our purpose is to analyze certain major accomplishments and to consider certain predicaments of sociological work today. We shall attempt to show how contemporary controversies are related to, and how they shed new light on, the classical problems in the field. Finally, we should like to determine whether these controversies suggest any tools that would be better suited to the study of basic problems of sociological analysis than these in use today.

Even a cursory glance at the state of sociological analysis (or at the state of affairs in anthropology and political science) reveals a somewhat paradoxical picture. Since the end of the Second World War, there has been unprecedented growth in all areas of sociological research—be it stratification, modernization, sociology of organization, or communications research. These subfields have been codified and systematized in an attempt to bring them into closer relation with theoretical frameworks. This consolidation in the 1950s and early 1960s was closely related to theoretical developments and controversies—almost conspicuously involving the structural-functional school which Talcott Parsons, Robert K. Merton, Kingsley Davis, and Edward Shils had done so much to develop. Not that the structural-functional school provided the reigning paradigm, but it was the focus of codification and of theoretical controversies in sociology and in other social sciences. These controversies, however, also caused the paradoxical situation in which far-reaching critical reexamination of the major assumptions of sociological analysis sometimes even denied the validity of the sociological enterprise.[2]

Because it would be impossible to survey all fields of sociological research, we shall focus on one central problem, and through its analysis we shall try to clarify the present situation and to understand its predicaments and possibilities. This problem is the analysis of the dynamics of societies and civ-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

ilizations, the nature of the processes of change within them, and the differences among them in these matters. We shall focus particularly on the problem of determining the conditions under which revolutionary change occurs—why have revolutions occurred only in some societies or civilizations, and only under certain historical conditions?

II

Macrosociological analysis in general, and the comparative study of societies and civilizations in particular, were major foci of sociological analysis beginning with the emergence of modern sociology in the early nineteenth century. This comparative interest was closely connected to the intellectual roots of modern sociology—to attempts to understand the nature of the emerging modern social order and how it differed from other types of social orders and civilizations.[3] Such macrosocietal analyses were focused on three major problems: first, the analysis of the general characteristics of "society" or social order; second, the comparative analysis of various types of societies and institutional complexes, with special emphasis on the uniqueness of the modern social order; and, third, the explanation of such variabilities in terms of some social or "natural" forces or mechanisms, which may also explain the trends in the developments of different societies and of human society in general.

The first great apogee of those studies came with the evolutionary approaches and the controversies that grew up around them. The great figures of evolutionary thought, such as Auguste Comte, Herbert Spencer, and Karl Marx, tried to indicate through a comparative analysis of societies, customs, or institutions, both the nature of modern society and what seemed to them to be the universal trend of its development through a series of stages. These stages were seen as leading to that end stage which would combine both higher levels of technology and social development and the extension of liberty, rationality, and progress.[4]

Even after the breakdown of these evolutionary syntheses—whether positivist or Marxist—macrosocietal comparative analysis with some evolutionary orientation constituted, at least until the early twenties of this century, one of the major critical theoretical concerns in sociology. Emile Durkheim, who was otherwise one of the severest critics of Positivist evolutionary sociology, retained in his distinction between mechanical and organic division of labor this comparative and evolutionary focus and interest.[5]

The combination of an evolutionary orientation with a strong emphasis on comparative studies can be found even in the most articulate "postevolutionary" approach to comparative studies—namely, the work of Max Weber.[6] Instead of using comparative analysis to illustrate a universal trend of development for all societies, Weber employed it to elucidate a particular trend predominant in one society or group of societies with a view to shedding light on similar or opposite trends in other societies. The most famous of his analyses was that of the economic orientations of the great world religions; he used it as a basis for the analysis of the specific religious constellations in Europe—especially the rise of the Protestant ethic, which, according to him, had given rise to the development of modern capitalism.[7]

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

But, as this work shows, Weber was concerned with two problems of comparative analysis in which some of the evolutionary perspectives still persisted. One was to understand the nature of modern societies (as distinct from premodern ones) and the degree to which the whole world might be moving in the same direction. The other problem, which can also be found to some extent in the works of Marx, Comte, and Spencer before him, was not only to understand the differences between modern industrial Western societies, on the one hand and traditional societies, on the other, but also to explain why modern society—with its capitalism, rational bureaucracy, and other features—developed only in Europe and whether the trend toward rationalization, bureaucratization, and demystification of the world constitutes a universal trend within all societies in general and within all modern societies in particular.

III

This interest in comparative macrosociological studies, in the dynamics of a variety of civilizations with a strong focus on the relations and contrast between modern and premodern Western and non-Western civilizations, reappeared after the Second World War in the studies of development and modernization that became a major focus of research in most of the social sciences—in sociology, political science, economics, and anthropology.[8] Its emphasis was on how to "develop" "underdeveloped" societies, and it engendered a whole spate of studies in all the social sciences, utilizing new analytical approaches and new methodological tools of inquiry—post-Keynesian and econometric studies in economics, in attitude studies, in survey research, in demographic and ecological research, and in analyses in sociology and political science. This work also linked up with some of the major theoretical developments in sociology and political science, especially with the "systemic" approaches to social and political life. The most influential of these systemic approaches was, as indicated above, the structural-functional approach developed in sociology by Talcott Parsons and taken up and further elaborated in political science by Gabriel Almond, David Easton, and others. These approaches defined societies or polities as systems, that is, as entities that have boundaries of their own to distinguish them from their environments, and that have mechanisms that maintain such boundaries and assure their continuity.

This combination of developments in sociological theory with research into the "Third World" reopened the major classical problems of sociological theory. These included the characteristics and internal dynamics of various types of societies, the nature of processes of change and of the conditions of stability of such societies, the process of transition from one type of society to another, and the extent to which such transition evinces a discernible universal evolutionary tendency from the simple to the complex.

The research that developed out of these concerns and dominated comparative studies in the social sciences in the 1950s and 1960s attempted first of all to identify the differences between traditional and modern societies.[9] These were defined in many ways by using sociodemographic indices such as urbanization, occupational structure, spread of media of communication, and the like. They were also defined in terms of structural differences—traditional societies

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

being characterized, to use Parsons's terminology, by particularistic and ascriptive criteria of role allocation, and modern societies by universalistic and achievement criteria. Distinction was made also between the "closed" personality type found in traditional societies and the more "open" one, with its capacity to empathize with a wide range of people and situations, that is concomitant with modernity. These differences between traditional and modern societies were couched in most of the studies in terms of the respective range of systemic problems with which they could cope or of the environments—both internal (social, cultural) and external (technological, economic)—which they could "master."

From this perspective, traditional societies were perceived as basically very restrictive and limited, whereas modern societies were seen as much more expansive and adaptable to a widening range of internal and external environments and problems. Special emphasis was given to the ability to cope with change in general and with economic development and industrialization in particular. The qualities of modern life, such as rationality, liberty, or progress, with which the classics of sociology were deeply concerned were here subsumed under these "expansive" "systemic" qualities of societies. Although not entirely neglected, those other qualities of the modern order were seen, or assumed, to follow naturally from the capacity to grow and to absorb change.

IV

Given the strong emphasis on development, it was not, however, sufficient to stress the differences between traditional and modern societies. No less important was the question of how the transition to modern could be effected. The model that guided this body of research implied that the conditions for the development of a viable, growth-sustaining, modern society, which comprised the continuous extension of the various sociodemographic and/or structural indexes listed above, necessitated the almost total destruction of all traditional elements. According to this view, the higher the structural specialization in a society and in its component organizations, the higher the rank of a society on various indexes of social mobilization demonstrating disintegration of traditional elements, the more likely the society to develop continuously and to be able to deal with new problems and social forces while expanding its institutional structure. This, in turn, would probably enable it to develop other qualities of modern societies such as rationality and a predilection to liberty and progress.

When scholars working within this paradigm addressed themselves to the problems of transition from a traditional to a modern setting, they developed two crucial assumptions that guided research in this field.[10] The first assumption was that the processes of modernization in the economic, political, and other institutional spheres tended to follow relatively similar patterns. The second assumption was that, once the institutional kernels of a modern system were established in one of these areas, they would lead to similar, irreversible structural and organizational developments in all social spheres and to sustained growth in the common evolutionary direction.

These assumptions were especially important in guiding the initial analysis

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

of "transitional" societies and of their variability, and it is through these very attempts to explain this variability that the initial model of modernization was gradually undermined. This is most apparent in the way the concept of stages was utilized in the study of transition from traditional to modern society and in the explanation of the variety and variability of traditional societies in terms of their ability to effect that transition.

Typologically, transitional societies, standing between traditional and modern societies on different indexes of "modernization," constituted a special "stage" in the development of human societies. However, in those studies this concept also acquired a dynamic connotation that stressed their historical transitionality, the inherent tendencies which pushed these societies in the direction of modernity. By virtue of this connotation the systemic qualities and the static, self-perpetuating tendencies of transitional societies were in the first studies of modernization subordinated to their presumed "dynamic" tendencies to develop in the direction of the end-stage of modernity, even though its was sometimes acknowledged that some societies may "halt" at some intermediate stage.

Accordingly, nonmodern societies were studied in terms of their relative resistance to internal or external forces of modernization. Even though there was a growing recognition of the possible diversity of transitional societies, it was still assumed that such diversity would virtually disappear at the end-stage of modernity. The theory of the convergence of industrial societies most eloquently propounded by Clark Kerr argued that ultimately all modern industrial systems would develop similar major institutional features.[11] Behind this loomed a conviction of the inevitability of progress toward modernity—be it political or industrial.

V

From the early and, especially, the mid sixties, the momentum of research and development on the world scene gave rise to far-reaching criticisms of these assumptions. These criticisms arose from a variety of vantage points, and they touched not only on the problems of development and modernization, but also on some very central questions of sociological analysis. Behind much of the debate there also loomed political and ideological differences, sometimes forcefully expressed. The two major foci of these criticisms were the alleged ahistoricity and Europocentricity of this initial model of modernization, and the closely connected tradition-modernity dichotomy.

The allegation of the ahistoricity and Western centricity of this model was developed in two concrete directions that bear directly on the problem of the dynamics of civilizations. One direction pointed to a reappraisal of the importance of historical continuity in shaping the contours of societal development. It was increasingly recognized that some of the differences in the structural and symbolic contours of modern societies and of their dynamics might be related to their respective historical traditions.

Perhaps one of the most important developments in this context was the concept of "patrimonialism" to describe the political regimes of some new states.[12] The use of the term "patrimonial" to describe many of the contempo-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

rary regimes pointed out the inadequacies of the central concepts and assumptions in the major studies of modernization, first by showing that many of these societies and states did not develop in the direction of modern European nation-states; second, by demonstrating that these societies were not necessarily a "transitional" phase along an inevitable path to this type of modernity; third, by indicating that there was nevertheless some internal "logic" in their development; and last, by emphasizing that at least part of this logic or pattern derived from the traditions of these societies themselves. Thus, these criticisms of the initial model of modernization stressed the importance of analyzing contemporary developments in various societies in terms of their "unfolding" of the traditional forces inherent in them, rather than their alleged movement toward a fixed end-stage.

The other—and, in a sense, opposite—direction that these criticisms took tended to emphasize the unique historical experience of the modern era. This approach, most clearly apparent in the works of many modern Marxists or semi-Marxists, stressed that the modernization process was not universal or inherent in the nature of every society. Rather, it was stressed that it represented a unique historical situation connected with the various aspects of European expansion, especially with the expansion of capitalism, and of the consequent establishment of a new international system composed of hegemonous and dependent societies. According to this view, the myopia of developmental theorists, if it is not ideological, results from their tendency to view societies in static isolation and from their failure to analyze the international pattern of relationships historically and contemporaneously. According to this theory of underdevelopment, the object of research is to study how both advanced development and underdevelopment can be explained by the structure and operation of one international economy.

Cutting across these two major directions, the claim that the initial model of modernization was ahistorical was reinforced in several additional ways. One was that this model, which stresses that the road to the envisaged end-stage of modernity and this end-stage itself are inevitably given in the concrete historical situations in which these potentially modernizing societies find themselves, precludes the possibility of choice. Thus, for instance, it was charged that the definition of "systemic" capacity of modern societies was very Europocentric and that it neglected the problem of choices among developmental goals. In this context it was claimed that because of its stress on development and economic expansion—derived from Western capitalistic experience and values—this model neglected the reconsideration of the qualitative aspects of modern societies, particularly of the possibility that among the various qualities of modern social order, such as liberty, equality, and growth, there may develop tensions and contradictions that necessitate making some difficult choices.

Ernest Gellner has succinctly summed up most of the criticism of this model for studies of modernization by pointing out the grave defects of assuming the "once-only, European-parochial, perpetual-progress" way of interpreting transition to modernity. Such an assumption tends to fuse and confuse several distinct sets of features, such as characteristics specific to the *first* such transition, those specific to the *European* transition, characteristics of *any* transition, and characteristics of a *completed* transition to modernity.[13]

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Contemporary Theoretical Controversies in Sociology

VI

Many criticisms of the initial model of modernization converged with the more general controversies around broad issues of sociological theory, most notably with those concerned with structural-functional theory and with the limits of its applicability to the analysis of social and historical change.

The major controversies in sociology gave rise to a variety of new models and approaches, such as the conflict model espoused by Ralf Dahrendorf and John Rex[14]; the exchange model espoused by George C. Homans and, to some extent, by Peter Blau[15]; and the symbolic structuralist one espoused by Claude Lévi-Strauss[16]; or to the reaffirmation and further development of older models—the symbolic-interactionist one, out of which ethnomethodology developed,[17] and the Marxist one.[18] The discussions around these various models and especially around the confrontation of these models with the structural-functional model constituted the focus of theoretical discussions and controversies in sociology in general, and in macrosociological analysis in particular.[19]

All these models claimed that the structural-functional model tended to minimize the autonomy of individuals and of subgroups, their interests and conflicts, and of the cultural or symbolic dimensions of human activity; and that it subordinated all of them to the organizational needs or exigencies of the social system.

Each model emphasized an aspect or component of social life presumably neglected by the structural-functional school. Thus, the exchange and conflict models emphasized the motives, interests, and conflicts arising in the society among various social actors, individuals, or groups; the symbolic-interactionist model emphasized that the normative definition of situations in which people interact is not simply given by the values of society, but is continuously constructed by people who interact in these situations. The structuralists and, to some degree, the Marxists emphasized some "deeper" symbolic or symbolic-structural dimensions of human life or of the social order as the keys to understanding the structuring and dynamics of societies.

Whatever the differences among these models with respect to the aspects of human activity that they emphasized, they all also claimed that the structural-functional model was unable to explain conflict and change. This, according to them, was because of its major assumptions, especially its strong emphasis on the close interrelation among various parts of a society; on boundary-maintaining mechanisms of social control; on a social consensus around common values and goals, and an implicit minimization of the importance of power and coercion as a means of social integration. Most of these models claimed that for all these reasons the structural-functional model was ahistorical and that within it the explanation of any concrete historical situation or phenomenon in terms of past influence and processes was relinquished in favor of a "static" or "circular" explanation theory with a very strong conservative bias.

Closely related to this contention was the unwillingness of most of these approaches to accept the "natural" givenness of any single institutional arrangement in terms of the organizational or systemic needs of the social system to

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

which it belonged. Any given institutional arrangement—be it the formal structure of a factory or a hospital, the division of labor in the family, the official definition of deviant behavior, the place of ritual in a given social setting, or the patterns of behavior that develop around that setting—were no longer examined only or mainly in terms of their contribution to the maintenance of any given group or society. Instead, the very setting up of such institutional arrangements was transposed from a given into a problem to be explained, or was "problematized"; it was asked what the forces were beyond the organizational needs of a society that could explain its major institutional arrangements.

The various models differed in their proposals as to how to cope with this problem, how to explain any concrete institutional order. One such approach, to be found especially in the individualistic and conflict models as as among the symbolic-interactionist ones, stressed that any such institutional order develops, is maintained, and is changed through a process of continuous interaction, negotiation, and struggle among those who participate in it. In this approach it was stressed that the explanation of any institutional arrangement has to be attempted in terms of power relations and negotiations, power struggles and conflicts, and the coalitions formed during such processes. Concomitantly, a strong emphasis was laid on the autonomy of any subsetting—subgroup, or subsystem—that could find expression in definitions of goals that differed from those of the broader organizational or institutional setting as well as on the "environments" within which the social setting operates—and above all on the international system for the analysis of "total" societies of macrosocietal orders.

The second, seemingly contradictory, approach was to be found especially among the structuralists and the Marxists. As indicated above, that approach tended to explain the nature of any given institutional order, and especially its dynamics, in terms of some principles of "deep" or "hidden" structure akin to those which, according to linguists such as Chomsky, provide the deep structure of language. In attempting to identify the principles of this structure, the structuralists stressed the importance of the symbolic dimensions of human activity, of some inherent rules of the human mind. In contrast, the Marxists stressed a combination of structural and symbolic dimensions—such as the dialectic between forces and relations of production and those among production, alienation, and class-consciousness—as providing the principles of the deep structures of societies to explain their crucial institutional features and dynamics.

VII

The very multiplicity of approaches to the most central issues indicates the need for a more differentiated, systematic analysis of some of the basic problems of sociological theory. Such an analysis should focus especially on how cultural orientations or traditions are interwoven into the institutional arrangements of a society; which aspects of the arrangement are influenced and shaped by central orientations—thus perhaps providing the principles of the deep structure of the society; who the major social actors are who act as the carriers of such orientations; what the social mechanisms are, that is, what the processes of social interaction and struggle are through which the carriers of cultural orientations

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

influence the shaping of these major aspects of institutional structure of societies, and what processes of change develop within them.

We shall attempt to illustrate the possibility of this more differentiated approach by coming back to an area which has been central in sociological analysis: the comparative dynamics of civilization, of which, as we have seen, the studies of modernization analyzed above have constituted the latest development. In this broad context we shall, as indicated above, focus on the problem of under what conditions revolutions or revolutionary change evolve and why they have evolved only in certain societies or civilizations and only under some historical conditions.

In attempting such an analysis we shall have to go beyond the earlier models of study of revolutions, which, just as were the models of modernization, were based on premises derived from the Western experience, especially from the European experience, of revolutionary transition to modernity. It is indeed one of the most interesting paradoxes of both classical and contemporary social analysis that the conception of revolutionary change as set forth mainly by Marx and the Marxists shares some basic concepts and assumptions with the ideologically presumably conservative evolutionists and neoevolutionists.[20]

All these approaches share a strong emphasis on the systemic nature of societies, on the close interrelation among the various parts of societies, on the crucial importance of social organization (even if they defined social organization in different ways) in structuring the activities of individuals, and on the dynamics of culture. Above all, they share some crucial assumptions about the nature of the transition among the various stages of social development—a concept which, as we have seen, was central to evolutionary and Marxist sociology alike. All these schools or approaches, as well as the contemporary neoevolutionary theories and neo-Marxist approaches, assumed that the transition from one stage to another involved a radical break with the past and concomitant changes in all spheres—political, social, economic—seen by both the evolutionists and the Marxists as the "real" and most significant changes in the development of societies, but particularly modern societies. Those who believed that such changes occur through a revolutionary process also assumed this process would take place through a combination of several types of collective action, such as rebellions, intellectual or religious heterodoxies, and central political struggle, as well as potential institution-building. This combination seemed to generate both great organizational capacities and a highly articulated ideology of social protest, especially of a utopian emancipatory ("leftist") nature, based on symbols of equality, progress, and freedom presumably leading to the creation of a new and better social order.[21]

However, these characteristics were not true of most of the social changes and transformations that took place throughout the major civilizations—be it the Islamic world, China, Japan, or even early Europe—in their "historical," "traditional" periods. No less important from the point of view of our analysis is the fact that many of the processes of change and transformation in modern Europe (such as Germany or Italy) or in other modernizing societies (such as Japan, where the Meiji Restoration generated far-reaching social transformation without a revolutionary-ideological symbolism) were not affected through such processes, nor have they always evinced such "totality" of change.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

In various civilizations and historical settings, a multiplicity of combinations of continuity and change have developed in the various spheres of institutional life. Moreover, in only very few societies have such changes taken place through the combination of rebellions, heterodoxies, central political struggle, and institution-building as envisaged in the "revolutionary" model. In each civilization, combinations formed among the various movements of change, and these combinations followed their own course toward change and transformation.

Such patterns of change, of the dynamics of civilizations, of the specific constellations of change that develop in various civilizations constitute a major challenge to sociological analysis and the major theoretical controversies in sociology analyzed above have indicated some of the problems which have to be coped with in order to attempt it. If such an analysis is possible, it will indicate that sociology is able to solve some of the theoretical problems indicated above and to go beyond the Western centricity of the earlier models of the dynamics of societies and to analyze systematically, on the basis of the analytical considerations presented above, variabilities of man's historical experience. It is this comparative analysis of the dynamics of civilizations that will be our major concern, and we shall focus it on the question of why revolutions have occurred only in some societies or civilizations, and only under certain historical conditions.

The Dynamics of Civilizations and Revolutions

VIII

Let us start with the simple, basic historical fact that the "real," "true" great revolutions have occurred in only very few societies—the Great Rebellion and the Glorious Revolution in England; the American Revolution; the French, Russian, and Chinese revolutions; possibly the revolt of the Netherlands; and the Turkish and the (North) Vietnamese revolutions—and that they all took place when their respective societies were at certain stages of historical transition.

It might be worthwhile, therefore, to see whether these societies, as well as the historical circumstances in which the revolutions took place in them, share some common characteristics. A closer look does indeed reveal, first of all, that all these societies belong to the imperial or imperial-feudal types of political systems, the most important of which in the history of mankind have been the Roman and Hellenistic, the Chinese, Byzantine, Russian, Abbasid, and Ottoman, and the major regimes of Western and Central Europe.[22]

Of special interest for our discussion is the fact that within these societies in their traditional settings distinct patterns of change developed which differed from those in other "traditional" or historical societies, especially the so-called patrimonial societies and most city-states and tribal federations. It was only in some of the Near Eastern—above all the ancient Israelite and Islamic—tribal federations and the great city-states of classical (Greek and Roman) antiquity that a different pattern of change, more similar to the imperial one, developed. The pattern of change that developed in these imperial and imperial-feudal systems also evinced some features relatively close to those envisaged as revolutionary change.[23]

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

The outstanding feature of the patterns of change in these societies was that a relatively high, although far from full, degree of coalescence between changes and restructuring of the major collectivities (political, religious, national) and of institutional frameworks (the economic, religious, and that of social stratification) tended to go hand in hand with a high degree of restructuring in the political system itself. These societies also developed strong connections among the various movements of protest and political struggle and a very high level of organization and articulation of the issues.[24]

Even within these imperial and imperial-feudal systems, the most frequent types of change were dynastic ones, sometimes affecting the boundaries of the polity, or, as in the Byzantine Empire, leading to their disappearance, but not producing far-reaching changes in their respective political structures. But beyond these types of change—although sometimes in close relation to them—other patterns developed in these empires and imperial-feudal societies, with a far greater degree of coalescence between the restructuring of the political regimes and other aspects of the macrosocietal order. In these empires, dynastic changes were often connected with changes in the composition of political elites, in the structure of broader social strata, in the political articulation, and in the impact of such changes on the central political institutions, especially with shifts in the relative predominance of some major groups in the political arena; and even with direct assumption or usurpation of political power and decision-making by newly developing groups, such as the military or some upper bureaucracy. Thus, such changes were often connected with the rise, growth, and strengthening, or the decline and weakening, of professional, cultural, and religious elites and institutions as against more ascriptive groups, with shifts in the strength of the monarch vs. the aristocracy and of the aristocracy vs. the urban groups and the free peasantry, or with shifts in the strength and independence of the bureaucracy. These changes were also often connected with changes in the principles of the political articulation of such groups, especially with the broadening or narrowing of their autonomous access to the center.

Concomitantly there developed in these societies a much closer connection among changes in the boundaries of the different components of the macrosocietal order than can be found in other types of traditional societies, such as the patrimonial city-states or tribal systems mentioned above. True enough, even in the imperial systems the actual boundaries of various ethnic, cultural, and political collectivities and institutional systems—the economic, that of social stratification, and to some degree the religious ones—tended to differ from one another. Many of the "great traditions" survived above all in the guise of churches, such as the Greek Orthodox, long after the disappearance of their respective empires. But with the sole exception of Islam, these great traditions tended to become, in such situations, less forceful and dynamic in sharp contrast to parallel developments in Buddhist societies or Hindu civilizations.[25]

A similar picture emerges with respect to the relation between the continuity of the economic frameworks and that of the political boundaries of the empires. Although the major economic entrepreneurs, particularly the commercial and "industrial" ones, had many "outside" connections and markets, a much larger portion of their activities was organized within the framework of the imperial boundaries and of their political peripheries than in most traditional patrimonial

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

city-states or tribal systems. Hence changes in the scope and structural principles of the economic systems or of strata formation tended to impinge directly on the political centers at the same time that far-reaching changes in the political regimes could affect the functioning of economic institutions and the structuralization of social hierarchies.

IX

In close relation to these patterns of change, the imperial and imperial-feudal systems tended to develop a high level of ideological articulation of political struggle, as well as a much closer connection between that struggle and the various movements of rebellion and protest.[26] This high level was apparent in the emergence of political leaders or entrepreneurs who attempted to represent the values or interests of broader groups and states, rather than confining themselves to "narrow" family organizational groups. Frequently, the rulers themselves, their representatives, or members of the ruling elite played leadership roles. In close relation to these types of political leadership more differentiated and autonomous types of political organization developed, in most of these regimes, in the form of cliques and semiorganized political parties. Naturally, the extent both of their organizational continuity and of their distinctive autonomy was much smaller than that of modern political parties.

These political systems also developed tendencies toward coalescence among the major types of protest, that is, between rebellions and heresies; between these and institution-building (primarily in the economic and cultural spheres by secondary elites); and between each of these and the more central political struggles and processes. In imperial and imperial-feudal regimes, some of these connections become more than just ad hoc coalitions. They gave rise to closer organizational and symbolic "merging" of these movements, often generating new symbolic and organizational institutional patterns. This last tendency was closely related to the high level of symbolic and ideological articulation of the political struggle mentioned above. Sometimes this struggle aimed at the redistribution of economic resources (especially of land and of debts) among the various social groups within the community. Sometimes it even developed into a struggle over the basic socioeconomic contours of the whole community; over the contents and patterns of participation in its center, and over the scope of participation of different social groups in the political life of the community—of what could be called, in more modern political parlance, the scope of political equality.

The pattern of change that developed in other traditional societies, especially in the so-called patrimonial societies and in most city-states and tribal federations, was very different. Although space does not allow for the analysis of this pattern in the same detail allotted to the imperial and imperial-feudal pattern, we may briefly point out some of its major characteristics: a relatively low level of coalescence between changes in the principles and boundaries of regimes and those of other collectivities and institutions, and the closely related tendency of segregation among various types of movements of protest and political struggle. A comparison between these two tendencies in traditional societies reveals a close connection among the degrees of convergence of various types of movements of protest, and between these protests and the central political struggle, on the one

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

hand, and the degree of coalescence among changes in the various components of the macrosocietal order on the other. The closer the connection among these movements, processes, and institutional innovations, the higher the degree of segregation between the various types of movements of protest and the political struggle, and the lower the degree of coalescence among changes in the various components of the macrosocietal order.

Clearly, even within traditional societies, especially in the classical city-states of Greek and Roman antiquity and in the Near Eastern, especially ancient Israelite and Islamic, tribal federations, there are some exceptions to this correlation between coalescence of movements of protest and convergence of changes in the various dimensions of the macrosocietal order. Moreover, within each of the ideal-typical patterns of change analyzed above was great variety, but we shall have to forgo its analysis here.

Instead we shall proceed to the next question: Is it possible to identify those social and cultural characteristics of the imperial and imperial-feudal societies, on the one hand, and the patrimonial ones, on the other, that can explain the development within those societies of varying patterns of change?

X

It is not easy to identify the conditions that account for the major differences in the patterns of change between imperial and patrimonial societies in terms of the variables often stressed in the sociological literature that we encountered in the analysis of modernization—namely, the type of social division of labor and degree of economic development, of structural differentiation, or of class relations or struggle. Rather, they can be identified in some other characteristics of these societies or civilizations that bring us closer to solving some of the problems facing sociological analysis as identified above.

Such characteristics can be found in some of the cultural orientations prevalent in these different types of societies. They can also be found in the combination of these cultural orientations with structural characteristics of societies that cut across similar levels of technology, on the one hand, or of structural differentiation or class composition, on the other. The most important of these characteristics, as we shall see in greater detail later on, are the structures of their centers and of center-periphery relations: the structuring of social hierarchies, of strata formation, and of the major collectivities.

It is very important to remember that most of these empires—with the partial exception of the Roman and Hellenistic ones—developed in close relation to some of the great civilizations or traditions in the history of mankind, such as the special Chinese blend of Confucianism, Taoism, and Buddhism; the Christian tradition in its variety; and the Islamic one. Most of these, with the partial exception of the late-comer, "Islam," emerged in the so-called Axial age, that is, in that age around the first millennium B.C. that saw the breakthrough of some of the great civilizations in various parts of the world, the respective "leaps" to history affected by various cultural innovations: the Hebrew prophets, the Greek philosophers, Jesus and the Apostles, Confucius, the codifiers of the Brahmin tradition, and the Buddha.[27]

Most of the traditions or civilizations within which these imperial or feudal-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

imperial societies developed shared several basic cultural orientations or codes that distinguished them from other civilizations that emerged in the same period. First, they were characterized by a conception of a high level of autonomy and distinction of the cosmic (religious) and mundane orders. Second, even though they shared this emphasis on the tension between the cosmic and mundane worlds with other civilizations, such as the Hindu and the Buddhist, unlike those, they included some kind of this-worldly activity, not only in the political, military, and cultural spheres, but especially in the European case also in the economic sphere, as a bridge between the transcendental-cosmic and the mundane worlds, or in Weber's terminology, as a focus of salvation. In addition, there developed within these civilizations a strong emphasis on the commitment of the different sectors of the population to the cosmic and social orders and a relatively autonomous access of at least some of the groups of these societies to the major attributes of some representative orders.

At the same time, these societies were also characterized by some specific properties of their centers, of center-periphery relations, and of the structuring of social hierarchies and collectivities.[28] The major characteristic of the center-periphery relations in the imperial, and to a large degree also in the imperial-feudal, societies was, first, a high level of distinctiveness of the center and of the perception of the center as a distinct symbolic and organizational unit, and, second, continuous attempts of the centers not only to extract resources from the periphery but also to permeate it and to reconstruct it. In addition, in many of these societies there also developed potential tendencies toward the impingement of at least part of the periphery on the center or centers.

The political and, to some degree, the cultural and religious centers in these societies were conceived as autonomous, self-contained foci of the charismatic elements of the sociopolitical, and often also of the cosmic cultural orders. These centers were the foci and loci of the various great traditions that developed in these societies, as distinct from local traditions, not only in content but also in the very symbolic and organizational structural characteristics. Unlike the patrimonial systems, the imperial systems worked on the assumption that the periphery could indeed have at least symbolic access to the center, largely contingent on some weakening of the social and cultural closeness and self-sufficiency of the periphery and on its developing some active orientation toward the social and cultural order represented by the center. This permeation of the periphery by the centers was discernible in their development of widespread channels of communication, which emphasized their symbolic and structural difference, and in the attempts of these centers to break through, even if only to a limited degree, the ascriptive ties of the groups on the periphery.

XI

Closely connected to this type of center-periphery relations is the development in these societies, despite some crucial differences that we cannot discuss here in detail, of some common characteristics in the structuring of social hierarchies and strata formation.[29] The first of these was the tendency within the imperial and the imperial-feudal societies for the boundaries of the major strata to resemble those of the respective political (and, in the imperial-feudal, the "civ-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

ilizational") frameworks. Unlike patrimonial societies, these frameworks and their symbols constituted basic referrents of the status symbols for at least the most active and powerful groups. Concomitantly, there developed in these societies some tendencies to autonomous articulation of widespread strata-consciousness, of relatively widespread status or class identity, as distinct from but related to the political and also to the cultural spheres. These broad tendencies found concrete expression in (a) the existence of at least some strata with autonomous access to social and cultural attributes which served as bases for the criteria of status, together with autonomous access to the centers of the society; (b) the development of some tendencies, especially among the more autonomous or central strata, to status-association among relatively diversified, even if similar, occupation of groups; (c) a relatively high degree—again mainly among the more autonomous or central strata—of country-wide strata consciousness and a less pronounced stratum formation of ethnic, religious, or regional elements; (d) tendencies toward some political articulation and expression of strata consciousness.

The combination of the political premises and characteristics of the imperial systems and the cultural orientations predominant in the great traditions to which they were connected also generated a certain dynamic in structuring the relations between the collectivities and the institutional complexes of their respective societies. First of all, a high degree of ideological symbolization grew up around the major religious, political, and even ethnical and national collectivities. Second, although in fact each such collectivity tended to develop a relatively high degree of autonomy as well as distinct boundaries, yet the various collectivities tended, in most of these civilizations, to constitute mutual referrents of each other (i. e., being a good "Hellene" was identified with citizenship in the Byzantine state, and vice versa). Within the societies the tendency also developed to stress a certain pattern of "ideal" coalescence among the collectivities, a tendency that was connected with attempts to arrange them in some hierarchical order—thus also providing a focus of continuous potential struggle among their respective carriers or representatives.

Thus in the imperial and imperial-feudal societies (and in the exceptional city-states and tribal federations of which limitation of space prevents our analysis here) the higher degree of symbolic distinction of the center and of strata formation was carried by a multiplicity of elites—both political and economic—as well as by representatives of the solidarity of different collectives having relatively autonomous bases and potentially autonomous access to the center and to each other. These elites impinged on the center and the periphery alike, shaping various movements on protest and of political activities and struggle which developed within them. Each of these elites could be a carrier of alternative conceptions of social order and could become a starting point for movements of protest or political struggle with a high level of organizational and symbolic articulation, by mobilizing the available free resources and organizing them into new directions.

In contrast, we find in patrimonial city-states and tribal regimes a combination of lack of structural distinction between center and periphery, a high degree of status segregation, weaker organizational and symbolic articulation of movements of rebellion and of conflicts, and minimal linkages between these

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

movements and the center. This combination was connected with a relatively low level of symbolic articulation of various collectivities as well as of secondary elites—whether functional ones, or of articulators of cultural models, or of the solidarity of different collectivities—and with a high degree of absorption of such institutional entrepreneurs into ascriptive groups with little symbolic and organizational autonomy, autonomous access to the center, or linkages among themselves.

XII

Had space permitted we could also have shown how coalitions among various types of elites and institutional entrepreneurs,[30] as well as the various cultural orientations of these entrepreneurs, explain some of the differences in the processes of change within the broad imperial and patrimonial patterns.

It is the structure of such entrepreneurial activities that provides the crucial link between cultural orientations and institutional structure. Differences in the structure of such entrepreneurs explain some of the relations between types of cultural orientations and institutional patterns in general, and different patterns of change in particular. The most important differences in the characteristics of such elites and entrepreneurs are the degree of their symbolic and institutional differentiation and autonomy and the relations between them and broader ascriptive groups.

To give only one illustration: Although political changes in Hindu civilization were mostly of the patrimonial pattern, the whole pattern of institutional change, the connection among changes in religious, economic, and cultural activities, was much more complex than those found in many "simple" ancient patrimonial systems. This difference is closely related to the structure of the Brahmin groups, their religious orientations, and their interrelations with the Indian caste system. Similar indications could also be given for the differences between the Chinese and Byzantine empires and for how the structure of their respective elite groups and cultural orientations were of crucial importance in the explanation of the uniqueness of the European experience—the first breakthrough to modernity. But all this must be left for later analysis.

But even at this stage some very important new questions arise. The first is the crucial question of whether it is the emergence of the various types of cultural orientations that gives rise to these elites or collectivities with different degrees of autonomy, or whether it is elites and collectivities with latent or manifest attributes of high or low levels of autonomy that create, generate, or select the orientations. Probably there is a process of continuous feedback through which groups or persons with certain potential structural tendencies select some orientations, and the institutionalization of the orientations selected then reinforced these structural tendencies. The systematic study of these processes of selection and feedback and of the mechanism of control through which the feedback is effected and institutionalized is one of the most important tasks for comparative sociological research. But the very posing of the question stresses that historical situations of change are potentially open, and that it is necessary to analyze how such situations become, as it were, closed in a certain direction. The importance of such analysis is brought out even more forcefully

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

if we go beyond the analysis of the dynamics of "traditional" civilizations to that of modern revolutions.

The fact of prevalence within the imperial or imperial-feudal societies of these cultural orientations and structural characteristics does not explain or assure the development of revolutions. Even in these societies the first "true," "pure" modern revolutions constituted a sort of mutation, an entirely new type of process of change. That change developed only in very specific sociohistorical circumstances, where relatively early stages of transition to modern settings entail a coincidence of three major aspects of the breakthrough from traditional to modern. These aspects are, first, transition from a traditional pattern of legitimation of political authority to a modern open one; second, the transition to an open class system rooted in, or connected with, a trend to a market economy in general and an industrial economy in particular; and third—and closely connected with the market economy—creation and/or incorporation of the respective societal units into a series of continuously changing international political, economic, and cultural systems.

In these types of historical conditions the potentialities for symbolically and organizationally linking protest movements, rebellions, heterodoxies, central political struggle, and institution-building could be fulfilled and focused on the reconstruction of the social order. It is in these circumstances that the potentialities of change through revolution become intensified and the special mutation of revolution *may* develop. It tends, however, to develop only—or mainly—in those societies that have been characterized by the coalescent pattern of change. In other societies similar processes of change may give rise to other results and patterns, but even in these societies it need not always develop, and here we are back to the problem of the openness or closedness of historical situations. Although the analysis of this problem is beyond the scope of this paper, it is important to realize that the problem arose out of the considerations presented in it.

XIII

We hope that our analysis of the dynamics of patterns of change in societies and civilizations, however preliminary and cursory, is of interest not only in its own right but also because it indicates some of the potentialities of new developments in sociological analysis described in the second section of this paper. It has stressed both the cultural and symbolic dimensions and the sociohistorical one. A further analysis—which for reasons of space was impossible here—would also have considered ecological factors. But our analysis has not attributed to any of these factors some ultimate reality, or seen in any of them the basic "causes" of the dynamics of societies and civilizations. Rather, it attempts to identify the social actors and institutional mechanisms through which these basic dimensions of human social and cultural activity are linked, in a relation of dialectic tension that provides the continuous focus of social and cultural dynamics. It has indicated the great variety of cultural orientations and of types of social organization, the ways in which cultural orientations and social organizations can become combined, and the mechanisms through which such connections are affected. It has stressed that, although these various elements of

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

social action exist in all societies, no specific combination of them or any specific pattern of change or development is the natural one, thus emphasizing also the relative openness of situations of change and of historical processes.

Such an approach can be applied in principle—and has already begun to be applied,[32]—to any field of sociological analysis, ranging from macrosocietal studies to those of organizations and small groups, to pattern of communication. Some very fruitful beginnings have already been made in several of these fields. But these are only beginnings. Much more work needs to be done.

References

[1]The recent reopening of the problem of sociological analysis are analyzed in detail in S. N. Eisenstadt and M. Curelaru, *The Form of Sociology—Paradigms and Crises* (New York: Wiley, 1976), Chs. 8-12.

[2]For the analysis of such extreme criticisms of sociological analysis, see Eisenstadt and Curelaru, *The Form of Sociology—Paradigms and Crises*, esp. Ch. 11.

[3]On the emergence of modern sociological analysis and the place of comparative analysis in it, see Eisenstadt and Curelaru, *The Form of Sociology—Paradigms and Crises*, esp. pp. 62-65; also R. Aron, *Main Currents in Sociological Thought*, 2 vols. (New York: Basic Books, 1966-1967); R. A. Nisbet, *Social Change and History: Aspects of the Western Theory of Development* (New York: Oxford University Press, 1964); and R. A. Nisbet, *Tradition and Revolt: Historical and Sociological Essays* (New York: Random House, 1968). On the place of comparative analysis in sociology, see also N. J. Smelser, *Comparative Methods in the Social Sciences* (Englewood Cliffs, N.J.: Prentice-Hall, 1976).

[4]On these scholars, see Aron, *Main Currents in Sociological Thought*, pp. 63-109; on Herbert Spencer see *Herbert Spencer on Social Evolution*, edited with an introduction by D. J. Peel (Chicago: Chicago University Press, 1972). See also R. Fletcher, *The Making of Sociology—A Study of Sociological Theory*, vol. 1 (London: Michael Joseph, 1971, pp. 165-196, 250-308). On Marx, see *Early Writings*, edited by T. B. Bottomore (New York: McGraw Hill, 1963), pp. 168-188, 189-194; and K. Marx, *Selected Writings in Sociology and Social Philosophy*, edited by T. B. Bottomore and M. Rubel (London: Penguin Books, 1965), pp. 175-185.

[5]See E. Durkheim, *The Division of Labor in Society* (Glencoe, Ill.: The Free Press, 1956), Book 1, Ch. 7; Book 3, Ch. 1; Aron, *Main Currents in Sociological Thought*, vol. 2, pp. 21-107; and *Emile Durkheim on Morality and Society*, edited with an introduction by R. N. Bellah (Chicago: University of Chicago Press, 1973).

[6]On Weber see O. Stammer (ed.), *Max Weber and Sociology Today* (Oxford: Blackwell, 1971); R. Bendix, *Max Weber: An Intellectual Portrait* (New York: Doubleday, 1960), Ch. 14; A. Giddens, *Politics and Sociology in the Thought of Max Weber* (London: Macmillan, 1972); W. Mommsen, *Max Weber—Gesellschaft, Politik und Geschichte* (Frankfurt am Main: Suhrkamp, 1974); and S. N. Eisenstadt (ed.), *Max Weber—On Charisma and Institution Building* (Chicago: University of Chicago Press, 1968).

[7]See S. N. Eisenstadt (ed.), *The Protestant Ethic and Modernization* (New York: Basic Books, 1968).

[8]These trends of research are summarized and analyzed in S. N. Eisenstadt, *Tradition, Change and Modernity* (New York: Wiley, 1973), especially Part 1, where also extensive bibliographical references are given.

[9]See, for instance, D. Lerner, *The Passing of Traditional Society* (Glencoe, Ill.: The Free Press, 1968); K. Deutsch, "Social Mobilization and Political Development," *American Political Science Review*, 55 (September 1961): 17-24; C. K. Black, *The Dynamics of Modernization—A Study in Comparative History* (New York, Harper & Row, 1966).

[10]For illustrations how these assumptions were worked out in research, see C. K. Black, *The Dynamics of Modernization*; W. W. Rostow, *The Stages of Economic Growth—A Non-Communist Manifesto* (Cambridge: Cambridge University Press, 1961); and, for a good summary, see D. Lerner, "Modernization—Social Aspects," in *International Encyclopedia of the Social Sciences* (New York: Grover, Collier & Macmillan, 1968), vol. 10, pp. 386-394. These assumptions have been subjected to a continuously growing criticism; see Eisenstadt, *Tradition, Change and Modernity*, Part 1.

[11]See C. Kerr, e.a., *Industrialism and Industrial Man* (Cambridge, Mass.: Harvard University Press, 1961); and C. Kerr, *Marshall, Marx and Modern Times: The Multi-Dimensional Society* (Cambridge: Cambridge University Press, 1969). For a criticism of these theories, see J. H. Goldthorpe, "Theories of Industrial Societies. Reflections on the Recrudescence of Historicism and the Future of Futorology," *European Journal of Sociology*, 12 (1971): 263-288.

[12]See G. Roth, "Personal Rulership, Patrimonialism and Empire Building in the New States,"

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

World Politics, 20 (1968); S. N. Eisenstadt, *Traditional Patrimonialism and Modern Neo-Patrimonialism*, Sage Professional Research Papers (Studies in Comparative Modernization), no. 90-003 (Beverly Hills and London: Sage Publications, 1973); and see also S. N. Eisenstadt (ed.), *Post-Traditional Societies* (New York, Norton, 1974).

[13]E. Gellner, *Thought and Change* (London: Weidenfeld & Nicholson, 1969), p. 139.

[14]R. Dahrendorf, *Class and Class Conflict in Industrial Society* (Stanford: Stanford University Press, 1959); R. Dahrendorf, *Essays in the Theory of Society* (London: Routledge & Kegan Paul, 1968); and J. Rex, *Key Problems in Sociological Theory* (London: Routledge & Kegan Paul: 1961).

[15]G. C. Homans, *Social Behavior: Its Elementary Forms* (New York: Harcourt Brace & World, 1961); P. M. Blau, *Exchange and Power in Social Life* (New York: Wiley, 1964).

[16]For Lévi-Strauss' symbolic structuralism, see C. Lévi-Strauss, *Structural Anthropology* (New York: Basic Books, 1963); *Totemism* (Boston: Beacon Press, 1967); *The Savage Mind* (London: Weidenfeld and Nicolson, 1966); and *The Elementary Structures of Kinship* (Boston: Beacon Press, 1969). Of the many works which provide interpretations and criticisms of Lévi-Strauss, the following are very instructive: E. Leach, *Lévi-Strauss* (London: Fontana/Collins, 1970); M. Glucksman, *Structuralist Analysis in Contemporary Social Thought* (London: Routledge & Kegan Paul, 1974), Chs. 1-3; I. Rossi (ed.), *The Unconscious in Culture: The Structuralism of Claude Lévi-Strauss in Perspective* (New York: Dutton, 1974).

[17]On symbolic interaction and ethnomethodology, see J. G. Manis and B. M. Meltzer (eds.), *Symbolic Interaction: A Reader in Social Psychology* (Boston: Allyn & Bacon, 1967); H. Garfinkel, *Studies in Ethnomethodology* (Englewood Cliffs, N.J.: Prentice-Hall, 1967); and A. Cicourel, *Cognitive Sociology* (Hardmondsworth, England: Penguin, 1973).

[18]On Neo-Marxist analysis, see J. Habermas, *Strukturwandel der Offentlichkeit* (Neuwied and Berlin: Luchterhand Verlag, 1962, 1965, 1968, 1969); *Sozialphilosophische Studien* (Neuwied: Luchterhand, 1963); *Technik und Wissenschaft als Ideologie* (Frankfurt am Main: Suhrkamp, 1968); *Knowledge and Human Interests* (Boston: Beacon Press, 1971); *Legitimationsprobleme in Spatkapitalismus* (Frankfurt am Main: Suhrkamp, 1973); and A. Touraine, *Production de la Société* (Paris: Seuil, 1973).

For a combined structuralist-Marxist approach, see L. Sebag, *Structuralisme et Marxisme* (Paris: Petite Bibliothèque Payot, 1964); M. Godelier, *Horizons, Trajets Marxistes en Anthropologie* (Paris: Maspero, 1973); C. Meillassou, *L'Anthropologie economique des Guoro de Côte d'Ivoire* (Paris: Mouton, 1974); many of the essays in the journal *Economy and Society*; and J. Habermas, *Zur Rekonstruktion des Historischen Materialismus* (Frankfurt am Main: Suhrkamp, 1976).

[19]See, among the many works of the structural-functional school, especially T. Parsons and E. Shils (eds.), *Toward a General Theory of Action* (Cambridge, Mass.: Harvard University Press, 1951); T. Parsons, *The Structure of Social Action* (New York: The Free Press, 1968); T. Parsons, *The Social System* (New York: The Free Press of Glencoe, (1951, 1964); T. Parsons and N. J. Smelser, *Economy and Society* (New York: The Free Press, (1956, 1965); and N. J. Smelser, *Social Change in the Industrial Revolution* (Chicago: University of Chicago Press, 1959). The controversies around this approach are analyzed in great detail in Eisenstadt and Curelaru, *The Form of Sociology*, esp. Chs. 8 and 9.

[20]The convergence between Marxist and neoevolutionary approaches can be found in K. Eder (ed.), *Die Enstehung von Klassengesellschaft* (Frankfurt am Main: Suhrkamp, 1973); C. Seyfarth and W. M. Sprondel (eds.), *Religion und Gesellschaftliche Entwicklung* (Frankfurt am Main: Suhrkamp, 1973). See also J. Habermas, *Zur Rekonstruktion des Historischen Materialismus*.

[21]On the image of revolution in modern social thought see M. Lasky, "The Birth of a Metaphor: On the Origins of Utopia and Revolution," *Encounter*, 34 (2–3) (1970); M. Lasky, *Utopia and Revolution* (Chicago: University of Chicago Press, 1976); K. Marx, *On Revolution*, arranged and edited with an introduction and new translations by S. K. Padover (New York: MacGraw-Hill, 1971); F. E. Manuel, "Toward a Psychological History of Utopias," in B. McLaughlin (ed.), *Studies in Social Movements* (New York: The Free Press, 1969), pp. 370–400; G. Landauer, *Die Revolution* (Frankfurt am Main: Rutten, 1912); A. T. Hatto, "Revolution: An Enquiry into the Usefulness of an Historical Term," *Mind*, 58 (1949): 495–517; A. T. Hatto, "The Semantics of 'Revolution,' " in P. J. Vatikotis (ed.), *Revolution in the Middle East* (London: George Allen and Unwin, 1972); E. Rosenstock-Hussy, *Revolution als Politischer Begriff in der Neuzert* (Breslau: 1931), Abhandlungen der Schlesischen Gesellschaft für Vaterländische Kultur, Geisteswischenshaftliche Reihe, 5 Heft; E. Griewank, *Der Neuzeitliche Revolutionsbegriff* (Frankfurt am Main: Suhrkamp, 1973).

Analyses of the modern image and concept of revolution can be most easily found in the many anthologies and surveys of the literature on social movements and revolutions which have recently appeared, and in which some of the most important studies of revolution—classical and modern alike—have been collected and analyzed, and in which materials bearing on our discussion will be found. Among such symposia and collections, of special interest are C. J. Friedrich (ed.), *Revolution: Yearbook of the American Society for Political and Legal Philosophy*, Nomos VIII (New York: Atherton Press, 1967); E. Kamenka (ed.), *A World in Revolution?* (Canberra: The Australian National University Press, 1970); J. Gerassi (ed.), *Towards Revolution: The Revolution Reader: Writings from Contempo-*

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

rary Revolutionary Leaders Throughout the World (London: Weidenfeld and Nicolson, 1971), 2 vols.; B. Mazlish, D. Kaledin, and B. Ralston (eds.), *Revolution: A Reader* (New York: Macmillan, 1971).

[22]The analysis of the imperial systems follows the one prescribed in S. N. Eisenstadt, *The Political Systems of Empires* (New York: The Free Press, 1963, 1969); and in the author's forthcoming book on *Revolutions, Social Change and Dynamics of Civilizations*. For some representative literature on some major imperial systems see G. Ostrogorsky, *History of the Byzantine State* (Oxford: Basil Blackwell, 1956); R. Ripes, *Russia under the Old Regime* (London: Weidenfeld and Nicholson, 1975); E. Balazs, *Chinese Civilization and Bureaucracy* (New Haven: Yale University Press, 1964); E. Reischauer and J. K. Fairbank, *East Asia, The Great Tradition* (Boston: Houghton Mifflin, 1960); B. Lewis, *The Arabs in History* (London: Hutchinson, 1950); P. M. Holt, A. K. Lambton, and B. Lewis (eds.), *The Cambridge History of Islam* (Cambridge: Cambridge University Press, 1970), 2 vols.; M. Bloch, *Feudal Society* (Chicago: University of Chicago Press, 1962), 2 vols.; O. Brunner, *Neue Wege der Verfassung und Sozialgeschichte* 2nd. ed. (Gottingen: Vandenhock & Ruprecht, 1968); C. Tilly (ed.), *The Formation of National States in Western Europe* (Princeton: Princeton University Press, 1975); J. O. Lindsay, "The Social Classes and the Foundation of the State," in J. O. Lindsay (ed.), *New Cambridge Modern History* (Cambridge: Cambridge University Press) vol. 7 (1957). For comparative cases of patrimonial societies, see Eisenstadt, *Traditional Patrimonialism and Modern Neo-Patrimonialism*, and the references cited therein.

[23]These patterns are analyzed in greater detail in Eisenstadt, *The Political Systems of Empires*.

[24]The patterns of political struggle are analyzed in Eisenstadt, *The Political Systems of Empires*, esp. Chs. 7–10; and in Eisenstadt, *Revolutions, Social Change and Dynamics of Civilizations*.

[25]On these patterns in Buddhist and Hindu societies, see S. J. Tambiah, *World Conqueror and World Renouncer* (Cambridge: Cambridge University Press, 1976); H. Bechert, *Buddhismus, Staat und Gesellschaft in den Landern des Theravada-Buddhismus* (Frankfurt am Main and Berlin: Alfred Metzger Verlag, 1966–1968), 4 vols.; M. Nash, G. Obehesekere, H. M. Ames, J. Ingersoll, D. E. Pfanner, J. C. Nash, M. Moerman, M. Ebihara, and N. Yalman, *Anthropological Studies in Theravada Buddhism* (Cultural Report Series no. 13, Southeast Asia Studies, Yale University, 1966); M. Singer and B. S. Cohn (eds.), *Structure and Change in Indian Society* (Chicago: Aldine Publishing Co., 1968); B. S. Cohn, *India: The Social Anthropology of a Civilization* (Englewood Cliffs, N.J.: Prentice-Hall, 1971); L. Dumont, *Religion, Politics and History in India* (The Hague: Mouton, 1970); L. Dumont, *Homo Hierarchicus* (Chicago: University of Chicago Press, 1970).

[26]See Eisenstadt, *The Political Systems of Empires*, especially Chs. 11 and 12, and the literature quoted there.

[27]See on these breakthroughs E. Voegelin, *Order and History* (Baton Rouge: Louisiana State University Press, 1954–1956), vols. 1–3; and the Spring 1975 issue of *Daedalus*, *Wisdom, Revelation and Uncertainty*.

[28]On the concept of center, see E. Shils, *Center and Periphery—Essays in Macro-Sociology* (Chicago, University of Chicago Press, esp. Chs. 1–3). The structure of center and center-periphery relations in these societies has been analyzed in Eisenstadt, *The Political Systems of Empires*, and especially the introduction to the 1969 paperback edition.

[29]The pattern of stratification in these societies is analyzed in greater detail in S. N. Eisenstadt, *Social Differentiation and Stratification* (Glencoe, Ill.: Scott, Foresman, 1972), esp. Ch. 6.

[30]On the concept of institutional enterpreneurs, see F. Barth, *The Role of Entrepreneur in Social Change in Northern Norway* (Bergen: Artok, 1963); S. N. Eisenstadt, *Essays on Comparative Institutions* (New York: Wiley, 1965), esp. Chs. 1 and 12; and S. N. Eisenstadt, "Societal Goals, Systemic Needs, Social Interaction and Individual Behavior: Some Tentative Explanation," in H. Turk and R. R. Simpson (eds.), *Institution and Social Exchange—The Sociologies of Talcott Parsons and George C. Homans* (Indianapolis and New York: Bobbs-Merrill, 1971), pp. 36–56.

Different types of institutional enterprises have been analyzed by Weber, above all in his work on the sociology of religion. See especially M. Weber, *From Max Weber, Essays in Sociology* (London: Routledge & Kegan Paul, 1947), Parts 3 and 4; *The Religion of India: The Sociology of Hinduism and Buddhism* (New York: The Free Press, 1958); and *The Religion of China: Confucianism and Taoism* (New York: Macmillan, 1964).

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

LESTER C. THUROW

Economics 1977

> It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way—in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.—Charles Dickens, *A Tale of Two Cities*

An Age of Imperialism

IF THE EARLY 1960s were the golden age of economics when economists were perceived as knowing all of the "right" answers, the late 1970s are the age of economic imperialism. The influence of economists has never been greater. Of the 16 Cabinet rank officials in the Carter administration, five have Ph.D.s in economics—only four are lawyers. Not so long ago lawyers had the corridors of public power to themselves. Yet the public esteem of economists has also never been lower. The only real burst of enthusiasm in the League of Women Voters' presidential primary panel on economics occurred when Senator Jackson heatedly told an economist questioner that it was time to stop listening to economists. In Johnny Carson's evening monologues economists have become ludicrous modern witch doctors unable to explain the economy's mysterious movements. At the same time economics has been recognized as a science with the establishment of a Nobel Prize.

Economists fill journals with articles on the economics of crime, marriage, suicide, and other social phenomena that were once regarded as the exclusive preserves of other social sciences.[1] They think of themselves as having a better analytical apparatus for studying these phenomena than other social scientists, and have had their imperialism accepted by much of the rest of the world. No self-respecting law or medical school can now exist without its economist. Economists regularly appear on the rosters of university departments of social work, engineering, urban and architectural studies, as well as traditional economics departments and business schools. Joint degree programs are springing up every-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

where with almost every other discipline. Graduate students in other departments and schools feel that they must take a substantial part of their coursework in economics. To protect their flanks against the inroads of economists, they must become economists. Economic forecasting has become a multimillion dollar business. The staffs of major antitrust cases now include as many economists as they do lawyers.

Yet the economy is widely perceived as out of control with economists having little of value to contribute. Double digit inflation and unemployment arrive without warning. The economy collapses. Economists neither predict these events nor seem to have cures for them. Traditional pieces of economic wisdom are proved wrong. Inflation and unemployment go up together. The stock market collapses as inflation accelerates. Personal savings rates go up as the value of the dollar goes down.

A Client Relationship

Part of the imperial success of economics is due to the fact that the profession understands that it has a client relationship with society. Economists have not been generally instrumental in shaping society's agenda, but they have been willing to work on that agenda—whatever it is. The profession is willing to change its research interests and focus of attention very rapidly in response to society's perceived needs and funding.

In the 1950s public interest in economic growth in underdeveloped countries and in staying ahead of the Russians led to an explosion of theoretical models of economic growth and empirical studies of the sources of growth.[2] Although growth had been traditionally explained in terms of expanding the capital and labor stocks, analysis indicated that a substantial fraction of the growth of the United States could not be ascribed to the growth of physical inputs. Output simply grew too rapidly to be explained by the growth of physical inputs. A substantial fraction of economic growth had to be explained in terms of improvements in the quality, as opposed to the quantity, of capital and labor. New pieces of equipment could undertake tasks that could not be undertaken by old pieces of equipment. Laborers learned new skills and managers learned to use the same units of capital and labor more efficiently with organizational improvements. As a result, the expansion of knowledge was seen to be a key ingredient in expanding the economy. Many of the post-Sputnik educational programs were in fact based on this conclusion. Education was the route to economic success for the individual and for the society.

Partly because of the importance of education in the growth process and partly because of society's interest in poverty and discrimination, the 1960s were marked by studies of the rates of return to various human skills and the extent to which low incomes were a function of discrimination or of inadequate skills.[3] In the process we learned that much of human skill acquisition did not occur in formal education and training, but in informal transmission of skills from one worker to another. Workers did not come into the labor market with fully formed skills, but acquired their skills while on the job. Skills might still be important, but they were not the skills acquired in formal education and training. Because formal education and training were traditional government

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

functions whereas most jobs were traditionally private, a problem immediately arose as to how governments were to alter the distribution of human skills when they did not directly influence the provision of human skills. To some extent the affirmative action programs of the late 1960s flowed out of this observation.

But something else also arose to confound conventional wisdom. Normally economists work with models where identical wages are paid for identical skills. Yet the real world did not seem to be marked by these homogeneous wages. For most skills there was a rather wide distribution of wage rates rather than a wage rate. One of the consequences of this was that standard economic variables (skills, IQ, hours of work, etc.) could explain only 20 to 30 percent of the variance in individual earnings.[4] Either economists were missing some key factor in determining earnings, or random error was much larger than had previously been suspected. If the latter was true, the world was much less deterministic and much more stochastic (see below) than had traditionally been believed.

As it became obvious in the late 1960s that skill augmentation programs were both expensive and relatively ineffective in altering the distribution of earnings, the public interest once again shifted. Welfare reform, minimum family incomes, and guaranteed jobs surfaced as major social and economic issues. A major part of the frustration in this area arose from the difference between what the public believed, or wanted to believe, and what economic analysis showed to be the case. In general it was cheaper to give many low skill individuals a lifetime income than it was to raise their earning capacities by the same amount. This led to the design of two variants of the negative income tax—President Nixon's family assistance plan and Presidential Candidate McGovern's demogrant plan. Self-sufficiency was obtainable, but only if the taxpayer was willing to put a very high price on "earned" as opposed to "transferred" income. As the presidential election of 1972 showed, however, taxpayers were not willing to pay the necessary price for self-sufficiency, but they also were unwilling to give up the ideal of individual self-sufficiency. Wanting the impossible meant that neither an efficient income transfer program nor an adequate skill augmentation program could be designed.

Partly because of frustration with "unsolvable" social problems and partly because of concern about the environment, the public's interest once again shifted in the 1970s and began to focus on the environment and the limits to growth. Economists were not instrumental in causing this shift to occur, but they responded by dusting off the concept of externalities, and small became beautiful.[5]

Whenever one person's actions have a direct nonmarket impact on another person's economic welfare, externalities exist. Individuals who cause the discharge of smoke and force those around them to breathe dirty air are creating negative externalities. Conversely, if I drain my land and make it cheaper for you to drain your land, I am creating positive externalities (disregarding environmental considerations). In our technically advanced and highly congested society, one group's actions much more frequently impact upon another group—airport noise. But our technology has also revealed long-standing externalities that we previously did not recognize—the cancer danger of asbestos fibers.

The standard economy solution to the problem of externalities is inter-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

nalization. Pollution externalities can, for example, be turned into market problems with "effluent" charges. Individuals and businesses are simply charged for their polluting activities until these activities are reduced to the desired level, until the funds raised from the charges are adequate to clean up the environment, or until the funds raised from the charges are adequate to compensate those who must suffer the effects of the pollution. Similarly, many safety problems can be turned into market problems with accident charges that are raised until they produce the desired level of safety or adequately compensate those being hurt. There are undoubtedly areas where such market solutions would be difficult to implement (cancer agents with long time delays might be one), but they could be used in many cases.

Yet despite the supposed influence of economists and our society's professed belief in market solutions, such solutions were resisted by everyone. Instead of internalizing externalities with "effluent" charges, society went off into direct regulation. But this merely breathed life into the field of industrial organization. This field had been moribund for a number of decades, but now the devices (antitrust laws and direct regulation) that had proved to be rather unsuccessful in dealing with the monopolies were to be applied to safety and the environment. Even though economists played a small role in designing these new rules and regulations, their major role has been a rear guard action arguing that the benefits caused by the rules and regulations were too small to justify the costs which they imposed.

When the unexpected energy crisis hit, energy economics almost instantly jumped from an area that was employing only a handful of economists to one that at times seems to be absorbing most of the profession. In a similar manner, the surge in agricultural prices breathed new life into agricultural economics. Intellectually stable industries are of little interest to economists. And society has little demand for their talents when prices are stable, output is growing on some smooth path, or economic discontent is slight.

In both the energy and food areas, substantial resources have gone into estimating better supply and demand relationships. At what prices will markets clear and what supplies are to be expected in the future? Both are critical questions. In the agricultural area a growing body of opinion points to the virtues of stabilization stockpiles that keep grain prices from fluctuating as widely as they now do. One of the reasons for doing so is to prevent agricultural prices from destabilizing industrial prices because of food's impact on wage demands.

Finally, there is the current perception that the economy is out of control. This perception began with the 1968 tax increase that did not slow the economy by as much as had been predicted, but it has been augmented by a series of unforeseen events ever since. The energy crisis, the food price explosion, the severe collapse of the economy in the winter of 1974–1975, and the intractable dilemma of inflation and unemployment all contributed to the perception. Although one can make a good argument that the economy was not nearly so much out of control as the public thought, the economics profession responded with new research on stochastic process and optimal control models where events are subject to random shocks. Rational expectation models arose to explain why public policies did not have the predicted effects.[6] (The policies had been anticipated and their effects were already embedded in economic actions.)

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

None of these developments, however, has led to a successful diagnosis or cure for the problem of inflation (see below).

To predict what economists will be working on ten years from now, one need merely predict what society will be interested in ten years from now. Perhaps this is one of the major differences between economics and the other social sciences. Society may wish to eliminate teenage street corner gangs, but this does not lead sociologists to write articles on the optimal techniques for eliminating such gangs. In contrast, society's interests seem to have a great impact on economics' intellectual path.

A Universal Methodology

Whatever the problem, mathematical models and their empirical analogue—econometric models—have now become the standard tools of economics. From their initial starting points in welfare economic and macromodeling, they have spread out to encompass everything.[7] No branch of economics, including economic history, is now free of their use.[8]

In the original Samuelsonian revolution, mathematical models were devices for exposing economic assumptions and the conclusions that spring from these assumptions. Mathematical modeling took on a life of its own (assumptions were made because they were mathematically tractable rather than because they were realistic), but there was also an inversion in the relationship between theory and data. Initially, econometric models were supposed to test whether the clearly specified theory could be statistically verified. Was the theory supported by the data?

In the end conclusive tests did not prove to be possible. Econometric models proved not to be up to the task. Equations and coefficients were not stable. Good historical equations proved to be poor predictors of the future. New data led to new coefficients. (In commercial macroeconomic models, equations change almost daily.) It also proved to be possible to build models that were equally good statistically from a number of different perspectives. Theories could not be accepted or rejected based on the data. Equations would not stand up over time. At any moment in time the models look solid and precise, but they are in fact elastic. The data simply were not powerful enough to test and to choose among theories.

As a result, econometrics shifted from being a tool for testing theories to being a tool for exhibiting theories. It became a descriptive language rather than a testing tool. Statistical models are built to show that particular theories are consistent with the data. But other theories are also consistent with the data and only occasionally can a theory be rejected because of the data. As a result good economic theory was stronger than the data—at least in the mind of the economists—and therefore it must be imposed on the data. What started out as being a technique for elevating data relative to theory ended up doing exactly the opposite.

Business investment functions are probably the best example of this phenomenon.[9] In classical economic theory, rising interest rates lead to less investment and falling interest rates lead to more investment. Much to the shock of initial investigators, econometric equations found either no effect or the exact

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

opposite to be true. Investment was not affected or went up when interest rates went up. Econometricians immediately went back to their computers to find an investment function where interest rates would be statistically significant and appear with the right sign. All such efforts failed.

As it became obvious, however, that econometric equations could not really test economic theories, another solution to the problem became intellectually legitimate. Econometric equations were designed in such a way that interest rates were mathematically forced to have the right sign. The equations did not test the theory, but they described what the world would look like if the theory were correct.

The theory of labor supply is another illustration of the change. In economic theory individuals are subject to income and substitution effects as they make their trade-offs between income (work) and leisure. If wages go up, the individual is being offered a larger bribe to undergo the disutility of work and will work more (the substitution effect.) But as his income goes up, the marginal value of income will fall relative to that of leisure and the individual will work less (the income effect.) Because of these simultaneous but opposing effects, economic theory cannot predict whether an increase or decrease in the wage rate will lead to more or less work. Economists need to know the empirical size of the income and substitution effects.

But econometric efforts to measure and identify income and substitution effects have led to a morass. Because every wage change embodies both an income effect and a substitution effect, the problem is to estimate independently either the income or substitution effect and then find the other by subtraction from the observed impact of wage changes on work behavior. Because individuals have nonwork sources of income, the standard procedure is to estimate how these independent sources of income affect work behavior and to then find the substitution effect by subtraction. The only problem is that independent sources of income seem to have a perverse effect on work.[10] Work effort goes up rather than down when nonwork income goes up. But given the estimating procedure, an income effect of the wrong sign will automatically lead to a substitution effect of the wrong sign. The smaller the bribe that people are offered to give up leisure, the more leisure they give up. Yet none of this makes any sense relative to economic theory. As a result macroeconomic models have given up on classical labor supply functions and make the labor force a function of job availabilities rather than wages.

Once again the result has been to elevate theory over data. Instead of starting with labor supply models derived from empirical estimation, models now start from mathematical representations of personal choice (utility functions) where work and leisure stand in the "correct" relationship to each other and where mathematics imposes a "correct" income and substitution effect.[11] These models are then used to predict the impact of some public policy, such as an increase in welfare payments, on labor supply. The models are used empirically but they are not empirical validations of the theory at all.

Although there may be nothing wrong with elevating theory over data as long as everyone realizes that this is what is occurring, it is such an easy way out that it tends to retard the development of economic theory. The theory is never challenged by the data and therefore never has to be rethought because it is

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

found to be empirically wanting. Economics has, for example, always treated work as an unpleasant disutility that one must be bribed to accept. All enjoyment—utility—springs from consumption rather than from production. To the economist, man is not a beaver who loves to work but a grasshopper who will work only if privation looms ahead. It is only now with demands for job enrichment that economic attention is beginning to swing over to look at producer's utility—the nonmonetary benefits that flow from work. Power, prestige, friendships, and feelings of accomplishment may all prove to be important parts of the reward system that have been left out by economists. Taken seriously they lead to a reversal of many standard economic conclusions. Although it is true, for example, that one can show that free trade maximizes consumption, it does not necessarily maximize producer's utility. If the French love to be farmers, it may be rational to protect French farmers with tariffs and quotas. What is lost in terms of consumption utility is more than gained in production utility.

The End of Newtonian Economics

The lack of stability in econometric equations has also led to another development. Econometrics has always paid lip service to the idea of stochastic (random) error terms, but it is only with the perceived breakdown in the economist's ability to control the economy that these have been taken seriously. Because real world stochastic shocks are not independent between quarters or even years, and because they simultaneously affect more than one economic variable, econometric equations have tended to exaggerate the deterministic as opposed to the stochastic element of economics. (In theory an equation with an R^2 of .99—not uncommon in time series analysis—says that only 1 percent of the variance in the dependent variable is stochastic and 99 percent of the variance is deterministic.) Stochastic disturbances are now taken much more seriously and thought to be much larger than has previously been the case. Hence the lack of stability.

Within economics there is a growth industry looking at risk and uncertainty, the nature of random stochastic processes, and their optimum control. The random walk in financial investments is the area where the most work has been done along this vein.[12] Instead of having a single rate of return on investments such as would occur in neoclassical economic theory, there is a distribution of ex post rates of return on investments that all had to be regarded as having identical ex ante expected rates of return. In a similar manner, economists are now discovering that there is not a single wage or a single price, but that there are distributions of prices or wages for identical skills or commodities.

There are a variety of pieces of evidence that have led economists to a random walk view of financial investments. Large financial firms (such as mutual funds) that invest a great deal of money and high-priced human talent in their investments have been unable to outperform the market averages or a random drawing of stocks. No one has been able to design a set of decision rules (when to buy and sell) that yields a greater than average rate of return. Tests indicate that stock prices quickly adjust to changes in public information (announcements of stock splits, dividend increases, etc.) so that no one can benefit from having public knowledge that is not already reflected in stock prices. And final-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

ly there is no serial correlation among stock prices over time. The price at any moment in time or its history cannot be used to predict future prices. When put together one has an argument that is similar to that advanced in quantum mechanics. Although it is possible to predict the center of mass of a distribution of stocks (atoms), it is not possible to predict the movement of any particular stock (atom). Given similar ex post distributions in wages and prices, random walk variants are likely to arise in many areas of economics other than financial investments.

Any discipline moves away from a deterministic view of the world with great reluctance. The deterministic nature of heavenly motion is the great example that led to modern science. No matter how complex events appear to be, they are in fact deterministic. It is only a matter of discovering the right perspective. Leaving this idea behind is always risky because it gives one an easy way out. What appears to be a stochastic process may in fact be deterministic if one only works harder and finds the right perspective. In any case economics, along with much of modern science, is being drawn in a direction where events are perceived to be much more stochastic and much less deterministic than had previously been thought.

Different Uses of Economic Models

Many of the disputes among economists and much of the imperial success of economics in the outside world spring from a failure to understand the different uses to which economic models can be put. First, they can be used to describe and organize events. Second, they can be used to understand what would happen in a controlled environment. Third, they can be used to predict what will happen in the real world. Fourth, they can be used to design policies to influence economic events. And fifth, they can be used as normative models to infer how a perfectly rational *Homo economicus* should act and how his activities should be organized.

Given a clever economist and completely elastic nonobservable concepts such as utility, psychic income, or long-run profit maximization, any activity can be described and organized as if it were the outcome of a free market maximization process. All consumption purchases represent utility maximization; all job choices represent income (psychic plus money) maximization; and all firm decisions represent long-run profit maximization. The ability to give such descriptions, however, does not mean that the model is an economic model in any of the other four senses.

Because economists cannot predict the elastic nonobservable variables, they cannot make economic predictions. Not understanding the process whereby these variables are generated, they do not understand economic actions. Without understanding, influence or control are impossible. Because these variables are typically part of the ultimate measures of economic value (the maximand), one cannot specify normative conclusions. Yet economists often talk as if they can do all of these things simply because they are able to build a descriptive model.

In its extreme form this claim springs from the following syllogism: (1) Free market-maximizing games are the most efficient economic games that man can

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

play. (2) More efficient economic games drive less efficient economic games out of business. (3) Therefore the actual economic game must be a free market-maximizing economic game. The only problem is to describe the game in the right way. Markets are as close to perfection as they can be, not because we have proved it, but because we know that it must be. Other than recommend that government get out of the way (there is a recent tendency to eliminate even this reservation), there is nothing that must be done to ensure maximum efficiency. As with all syllogisms, the real problem is to understand the conditions where it is likely to be true (commodity markets, for example) and the conditions where it is less likely or even unlikely to be true (natural monopolies, for example).

Although the vision of perfect competitive markets holds a powerful influence on economics, there has at the same time arisen a small subcurrent that is essentially its polar opposite—radical political economics. This strand of economics grew out of the political turmoil of the Viet Nam War and is an Americanized version of Marxist economics. In radical political economics attention focuses almost entirely on the existence and exercise of economic power. Monopolies and oligopolies abound and are tightly integrated with other forms of power. Economic power buys political power; political power defends economic power. Schools produce the docile work habits the capitalists desire and serve to provide a facade of meritocracy for what is in fact a perpetual class of power holders.[13] Because economic power is tightly integrated with the social structure, radical political economists believe that all attempts to influence or control this power are ineffectual. The only answer is a revolution which will overthrow the entire system. But because a revolution is necessary, a detailed understanding of how the current system works is not necessary. Radical political economists know that economic power exists to exploit the working class, but they are unable to predict or understand the precise instruments it will use to do this.

There is also a fundamental confusion between economic understanding and the ability to make economic predictions. Most physical sciences understand rather than predict. To understand a phenomenon is to be able to predict the outcome of a laboratory or controlled environment experiment where "other" variables are held constant and stochastic processes are limited. With the single exception of heavenly motion, physical scientists predict outcomes in controlled environments. They are no better, for example, at predicting the real world's meteorological phenomena than economists are at predicting the real world's economic events. To judge economics in terms of its predictive record is to judge it in a way that no other science is judged.

Although the use of economic models to design policies to influence economic events springs from our understanding of economic processes, economists have not placed enough emphasis on the distinction between control and influence. For the same reason that economic events are not predictable, they are not completely controllable. At any moment in time there will be other unpredicted influences and stochastic shocks will occur. The 1968 tax increase influenced the economy—it grew more slowly than it would have without the tax increase—but the economy expanded faster than was desired or predicted. Policies exist to control the economy partially, but policies do not exist to control the economy completely.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

If one looks at the different uses to which economic models can be put, it is clear that economics stands apart from all other social or physical sciences in one major respect. It has a normative model of how *Homo economicus* ought to act. The concept of a rational economic man is not just a concept of how men do act, but a concept of how they ought to act. Economists are preachers as well as teachers or scholars. Business schools are to some extent an embodiment of the economist's normative model. Business managers are convinced that they ought to profit maximize and are taught how to profit maximize. Economists are in a peculiar position in that if they are completely successful in their preaching function they will be completely successful in their scientific function because they will have created the world they seek to describe, predict, understand, and influence.

A Riddle Wrapped in an Enigma Inside a Conundrum—Inflation

The popular press has decided that economists do not understand the processes that produce inflation and unemployment, but it is half wrong. Unemployment is well understood. Equations predicting the growth of the labor force may fail to capture quarterly surges or ebbs in the growth of the labor force, but over relatively short periods they are some of the best of econometric equations. The relationship between employment and the real growth of GNP (sometimes known as Okun's law) is similarly stable. Employment may grow faster than would have been predicted from the real rate of growth of GNP for short periods (as it did in the last quarter of 1975 and the first quarter of 1976), but there always follows a catch-up period where the standard relationship reestablishes itself (as it did during the last three quarters of 1976). The unpredicted surge in unemployment in the winter of 1974–1975 occurred because of a failure to predict the magnitude of the decline in real GNP, not because of a shift in the relationship between employment and real growth. Unemployment becomes an economic problem because policy-makers see it as their only device for controlling inflation. But this is precisely what economists do not understand about unemployment—its impact on inflation.

Economists do not have models for understanding, predicting, or controlling inflation, but they do have at least two models for describing inflation. The first description factors inflation into its different sources and the second description focuses on necessary relationships among monetary variables.

From the perspective of the first descriptive technique, the double-digit inflation of 1974–1975 is factored into that part which can be traced to OPEC's tripling of the price of energy, the world-wide bad weather leading to escalating grain prices, capacity shortages in various raw material industries arising from simultaneous economic booms in most of the industrial world, and to dollar devaluations and rising prices for imports and exports. Although economists will admit that they should have foreseen the results of the last two factors (and any residual which is not explained by these four factors), they are not willing to take responsibility for that part of inflation which came from the first two. Both of these are events (political and meteorological) that economists could not have been expected to predict, much less control. They are just real-world examples of why the economy is stochastic rather than deterministic. Unfortunate acci-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

dents are a part of the economic process and no one can expect economists to have a political theory of cartel formation or a meteorological theory of bad weather.

The price effects of the dollar devaluation are something that should have been predicted, but they are not something that was controllable or preventable. If a system of fixed exchange rates had allowed the dollar to become overvalued, the dollar at some point in time would have to be devalued, and whenever this happened import prices and export demands would go up. If rising export demands lead to rising prices for those items exported, then the American consumer would pay higher prices for both those things imported and those things exported. But even leaving these three sources of inflation aside as unpreventable still leaves some significant fraction of the inflationary problem to be explained. Why did businessmen not see the raw material capacity shortages that were going to develop and build the necessary capacity? Why did wages rise rapidly in a period when unemployment was high? Currently prices are rising 5 to 6 percent per year, yet most of the impact of these three unpreventable factors is behind us.

The monetary description of inflation received renewed life as monetary economics (led by Milton Friedman) reestablished itself after the Keynesian revolution.[14] This reestablishment occurred for a number of reasons. The extreme Keynesians—but not Keynes himself—clearly made overly ambitious claims that were going to be deflated. The supply of money was important. Econometric models that left out monetary variables did not work as well as those that included them. Some sectors of the economy, such as construction, simply could not be explained without the use of monetary variables. In addition, the forced savings of World War II had built up an enormous degree of liquidity that would eventually disappear. When it did disappear, liquidity and the supply of credit again became an important variable. Just as any business or household, the economy had a balance sheet as well as an income statement. Changes in the stocks affect the flows (the stock of wealth affects the flow of consumption) and the desired stocks in part depend upon monetary variables—interest rates and credit availability.

The debate about the relative role of monetary policies (interest rates and the supply of money) and fiscal policies (aggregate tax and expenditure decisions) is more heated than normal because the debate tends to occur along political lines. The fiscalists tend to be liberal whereas the money supply, but not interest rate, monetarists tend to be conservative. The political debate revolves around the question of whether government expenditures should grow or contract, but it tends to get fought out in monetarism versus fiscalism because there is always an argument as to whether rising government expenditures will, or will not, return the economy to full employment in recessionary periods.

The basic equation used in a monetary description of inflation (see equation (1)) is true by definition. The real question is the extent to which this descriptive equation allows one to predict, understand, or influence inflation. There are two basic issues. First, although the supply of money (M) is determined by government monetary authorities, MV is not under their control unless they can predict or control the velocity of money (V) as well as the supply of money. Second, there is the problem of what fraction of any change in MV shows up in

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

changes in the price level (P) and what fraction shows up in changes in the real gross national product (T).

$$MV \equiv PT \tag{1}$$

where: M = stock of money
V = velocity of money
P = price level
T = real GNP or transactions

To be useful for something besides description the quantity theory of money—equation (1)—must be supplemented with a subsidiary equation explaining the velocity of money. The critical questions are the extent to which the velocity equation is stable and the size of the stochastic errors relative to the deterministic part of this equation. If the velocity of money equation is unstable or subject to large stochastic errors, it is not possible to predict or control MV even through the government controls the supply of money (M). The descriptive theory is valid but of little use. Part of the dispute between interest rate monetarists and money supply monetarists revolves around the extent to which the velocity of money is exogenous and the extent to which it depends upon the rate of interest.[15] The current hope is that microanalysis of the demand for money will yield a stable macrovelocity equation that we do not now have. But this is not at the moment a reality. The necessary stable velocity equation does not now exist.

Economic understanding of the relationship between P and T—what causes prices to go up and what causes real growth—has undergone a series of transformations as theories have been postulated and rejected. Classical economists thought that increases in monetary stimulus (a rising MV) would raise the real GNP (T) until full employment was reached. At that point a discontinuity would be reached and further stimulus would lead to an increase in the price level (P). Keynes believed the same thing with respect to fiscal stimulus. Prices would not start to rise until the economy had reached full employment. Problems arose because empirically prices start to rise long before the real GNP has reached what was conventionally believed to be full employment. (One semantic solution is to redefine full employment to be that point where prices start to rise, and this is to some extent happening at the moment.)

The observation that prices start to rise long before the real GNP reaches full employment levels led to the Phillips curve hypothesis.[16] According to this hypothesis, there was a stable relationship between the quantity of idle resources (as measured by the unemployment rate) and the division of changes in PT between price increases and increases in the real GNP. As idle resources fell, any given quantum of fiscal or monetary stimulus would lead to larger and larger increases in prices relative to those in the real GNP. Instead of a sudden shift from real growth to price inflation, there was a range of deteriorating tradeoffs that was usually explained in terms of more and more industries reaching capacity constraints or rising oligopoly power as idle resources shrank.

The Phillips curve hypothesis has only one problem. No one has ever been able to find stable price equations in U.S. data. Price and wage equations are the poorest equations in econometrics. The Phillips curve idea has had a long half-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

life because it must be true at the extremes. If there were no idle resources, all stimulus must lead to increases in prices alone. There must also be a level of idle resources so high that prices would fall. If all types or resources were in surplus, monetary or fiscal stimulus must lead to real growth. If the notion is true at the extremes, why should it not be true over the mid-ranges?

The currently popular natural rate of unemployment hypothesis returns to the idea of a discontinuity or switching point, but the discontinuity is not necessarily at what would conventionally be considered full employment.[17] In this view the key element in inflation is labor's demand for real wage increases. Prices are determined by wages. At each level of unemployment labor demands some real wage increase and asks for some nominal wage increase which it thinks will generate the desired real gain. At the natural rate of unemployment labor will ask for some nominal money wage increase. This nominal wage increase may be greater than the rate of growth of productivity. If so, it will lead to inflation, but when labor subtracts the rate of inflation from its nominal wage increase it will find that it has achieved exactly the real wage increase which it desired. As a result it will ask for the same nominal wage increase in the next wage round. It has no incentive to do otherwise. The resulting rate of inflation may be high or low, but it will be stable.

If economic policymakers push the rate of unemployment below the natural rate, labor will demand higher real wage gains and ask for higher nominal wage increases. But these higher nominal wage gains can lead only to more inflation because the real gain is limited by the growth of productivity. The real wage gain, when inflation is subtracted from the nominal gain, will be below that now demanded. As a result labor will demand even higher nominal wage increases in the next round. But each higher nominal wage increase is offset by a higher rate of inflation. Inflation accelerates round after round. Conversely, if unemployment is above the natural rate, prices fall. Because of the higher unemployment rate labor reduces its real wage demands below the level determined by productivity gains. With labor productivity growing faster than wages, unit production costs fall, leading to price reductions. But with falling prices labor finds that its real wage gain is greater than what it demanded and so it reduces its nominal wage demands in the next round. As a result prices continually fall. Thus the real GNP (T) is set by the natural rate of unemployment and all attempts to push unemployment above or below this point will show up only in changes in prices (P) in the long run.

The natural rate of unemployment hypothesis generates a number of unanswered questions. Why should labor be able to impose its demands for nominal wage increases? Only 25 percent of the labor force is unionized. Labor unions may have the monopoly power to impose above equilibrium wages in the union sector of the economy, but this will lead only to a reduction in employment in this sector and an increase in the labor supply and a reduction in wages in the nonunion sector. But if unions cannot be used to explain inflexible demands for real wage increases, where do these demands come from? Even ignoring this problem, what determines labor's nominal and real demands? Are labor's preferences for nominal and real wage gains stable? If not, the natural rate (or the equation describing it) will not be stable. What public policies exist to alter labor preferences and the natural rate? How fast does inflation accelerate and

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

decelerate if unemployment is above or below the critical point? Are the time lags stable or symmetric? Where is the natural rate? Can it be found? What stable rate of inflation is associated with it? Any stable rate is in accordance with the hypothesis.

To the real pessimist increases in fiscal or monetary stimulus primarily affect P (produce inflation) whereas reductions in fiscal or monetary stimulus primarily affect (produce unemployment). This asymmetric relationship in prices and wages is one of the fundamental recurring mysteries of economics. Although prices and wages are not completely rigid in a downward direction (witness the recent auto rebates and the construction unions agreeing to work at 75 percent of their normal pay on rehabilitation projects), there clearly is an asymmetric relationship. Prices and wages easily rise to clear markets, but they do not fall to clear markets with the same ease and speed. Why? Even the Great Depression with its 25 percent unemployment rates and 10-year duration managed to reduce prices by only 20 percent.

One explanation is to argue that prices are related to costs—ultimately wages. Wages are in turn rigidly downward because of an "implicit contract" between the worker and the employee not to reduce nominal wages. But this just raises the question as to why this implicit contract was made and how it is enforced.

The Missing Link—A Theory of Oligopolistic Behavior

To have a useful macroeconomic theory of inflation it is necessary to have a good microeconomic theory of price determination. Economics has a supply and demand theory of price determination for competitive markets, but larger and larger fractions of the GNP are being produced in sectors (medicine, government, etc.) that are clearly not competitive profit-maximizing sectors. Many other industries have just a few competitors, and the GNP has become increasingly concentrated among the largest firms in the last two decades.

To explain inflation economics needs an empirically useful theory of oligopolistic behavior. How do very large firms, unions, and nonprofit agencies act when they have few competitors or when they are not simple income or profit maximizers? What causes them to change their prices? How do they decide on the right mix between price competition and competition on other dimensions—product differentiation, service, etc.? If economics does not have a theory of price determination in oligopolistic situations or if prices are indeterminant in such situations, then it cannot have a useful theory of inflation.

With the publication of *The Theory of Monopolistic Competition*, *Competition Among the Few*, *The Economics of Imperfect Competition*, and *Theory of Games and Economic Behavior*, there was every reason to believe a few decades ago that progress could be made in developing a useful theory of oligopolistic behavior.[18] Some things were learned (oligopolies will tend to have excess physical capacity) in all of this effort, but it did not lead to a deterministic pricing theory such as that which exists for competitive firms or individuals. Realistic sets of assumptions always seemed to lead to indeterminant results. Deterministic solutions required counterfactual sets of assumptions. The world is not a two-person zero-sum game, and no oligopolist expects competitors to hold their prices or

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

quantities fixed in response to his moves. Yet these are the assumptions that lead to deterministic results. Without a theory of oligopolistic behavior it is impossible to predict, understand, or influence oligopolistic behavior or even to build a normative model of how oligopolists should act. They can be described as long-run profit, wage, or size maximizers, but this has no operational significance. Without a theory of oligopolistic behavior there is no microeconomic theory of price determination.

One solution would be to create markets that we do understand—competitive markets. But for a variety of reasons, economists' interest in creating competitive markets has waned. Those who believe the earlier syllogism believe that markets are in fact much more competitive than they appear to be. Radical economists believe that all partial efforts to increase competition—reduce economic power—will fail. Using antitrust laws to slightly enlarge the number of oligopolistic competitors does not seem worth the effort. Regulating oligopolies is popular with the public but tends to be seen by the economics profession as producing more inefficiency than it eliminates. If one does not understand oligopolistic behavior, how can one know that benefits of eliminating it exceed the costs? As Galbraith has argued, oligopolies may lead to types of de facto planning which society needs to be dynamically efficient.[19] Or, they may cause or lead to the acceptance of more technical change that would otherwise be the case.

Economics needs a theory of oligopolistic behavior and especially a theory of oligopolistic price determination. There are economists who deny the meaningful existence of oligopolies; there are economists who glorify their usefulness; there are economists who blame them for almost everything. But there are no economists who understand them.

Conclusions

Although society will undoubtedly generate enough problems to keep economists busy applying their existing intellectual apparatus, the profession's internal development is uncertain. It proceeds in a manner similar to that of gold mining. Some great, or lucky, prospector strikes a vein of high grade ore in the form of a new paradigm, technique, or vision. Given the knowledge of the existence and general location of the new vein of ore, ordinary miners go to work to mine much of the actual gold. Eventually, however, the vein plays out and the miners must work harder and harder to get less and less. The intellectual rewards of further and further research along that line get smaller and smaller.

In the last half century the great intellectual gold strikes have been the national income accounts, Keynesian economics (a development that also allowed monetary economics to be rebuilt or rediscovered), mathematical economics, and econometrics. Lesser gold strikes include the analysis of oligopoly, growth, human capital, and the random walk. Some of these strikes are still being worked in the last half of the 1970s, but most of them seem to have reached relatively low grade ore. The last decade has not witnessed a major or even a lesser gold strike. To rejuvenate its internal intellectual growth the profession

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

needs a gold strike, but as with all actual gold strikes, no one knows where, or if, it will occur.

References

The author would like to thank Paul Samuelson, Robert Solow, Peter Diamond, and Edward Kuh for their aid in preparing this paper. They are not responsible for any of the conclusions. The references in this paper are meant to be illustrative of the points being made. They are by no means exhaustive or necessarily the most important papers in the area.

[1]Gary S. Becker, "A Theory of Marriage: Part I," *Journal of Political Economy*, July-August 1973, p. 813.

[2]Robert Solow, "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 39 (1957); and Edward F. Denison, *The Sources of Economic Growth in the United States and the Alternatives Before Us* (New York; Committee for Economic Development, 1962).

[3]Richard Eckhaus, *Estimating the Returns to Education* (New York: Carnegie Commission, 1973); Gary Becker, *The Economics of Discrimination* (Chicago: University of Chicago Press, 1957); and Anthony Pascal, *Racial Discrimination in Economic Life* (Lexington, Mass.: D.C. Heath, Lexington Books, 1972).

[4]This point was first emphasized by a noneconomist, Christopher Jencks, in *Inequality* (New York: Basic Books, 1972).

[5]E. F. Schumacher, *Small Is Beautiful* (New York: Harper & Row, 1973).

[6]Robert S. Pindyck, "Optimal Policies for Economic Stabilization," *Econometrica*, May 1973; William Poole, "Rational Expectations in the Macro Model," *Brookings Papers on Economic Activity*, no. 2 (1976).

[7]Paul Samuelson, *Foundations of Economic Analysis* (Cambridge, Mass.: Harvard University Press, 1947); and Lawrence Klein and A. Goldberger, *An Econometric Model of the United States* (Amsterdam: North-Holland, 1955).

[8]Robert W. Fogel and Stanley L. Engerman, *Time on the Cross* (Boston: Little, Brown, 1974).

[9]J. Meyer and E. Kuh, *The Investment Decision* (Cambridge, Mass.: Harvard University Press, 1957); Dale W. Jorgenson, "Anticipation and Investment Behavior," in J. S. Dusenberry, E. Kuh, Fromm, and L. Klein (eds.), *Brookings Quarterly Econometric Model of the United States* (Chicago: Rand McNally, 1965); and Robert Eisner and M. I. Nadiri, "Investment Behavior and Neoclassical Theory," *The Review of Economics and Statistics*, August 1968.

[10]Marvin Kosters, "Effects of Income Tax on Labor Supply," in A. L. Harberger and M. J. Bailey (eds.), *The Taxation of Income from Capital* (Washington, D.C.: Brookings Institution, 1969).

[11]Martin Feldstein, "On the Optimal Progressivity of the Income Tax," Harvard Working Paper no. 309.

[12]An excellent survey article is Eugene F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Works," *Journal of Finance*, May 1970.

[13]S. Bowles and H. Gintes, "IQ in the U.S. Class Structure," *Social Policy*, January-February 1973.

[14]Milton Friedman and Anna Schwartz, *A Monetary History of the United States, 1867–1960* (Princeton: Princeton University Press, 1963).

[15]Franco Modigliani and Albert Ando, "The Relative Stability of Monetary Velocity and the Investment Multiplier," *American Economic Review*, September 1965. Reply by M. Friedman and D. Meiselman and rejoinder, same issue.

[16]A. W. Phillips, "The Relationship Between Unemployment and the Rate of Change of Money Wages in the UK, 1861–1957," *Economica*, November 1958.

[17]M. Friedman, Presidential Address, *American Economic Review*, March 1968.

[18]Joan Robinson, *The Economics of Imperfect Competition* (London: Macmillan, 1933); Edward Chamberlin, *The Theory of Monopolistic Competition* (Cambridge, Mass.: Harvard University Press, 1933); William Fellner, *Competition Among the Few* (New York: Knopf, 1949); and John Von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (New York: Wiley, 1944).

[19]J. K. Galbraith, *American Capitalism* (Boston: Houghton Mifflin, 1952); and *The New Industrial State* (Boston: Houghton Mifflin, 1967).

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

JONATHAN CULLER

In Pursuit of Signs

IT IS A COMMONPLACE of historiography that decisive events are difficult to perceive, except retrospectively. It is the future which will promote incidents of our own time to the status of events and which will enable those events to take their places in the causal sequences we are pleased to call history. Much the same can be said of events in the history of scholarship. Generally, it is not until years afterward, perhaps not until an influential movement has run its course or changed its direction, that one can identify as crucial the events which led to its foundation or determined its development. And therefore the observer of contemporary scholarship who wishes to make projects, articles, and alliances into the stuff of history is compelled to imagine a future from whose proleptic vantage point he can construct causal sequences and narrate the "real history" of his own time.

This kind of temporal projection, however fanciful it may seem, is the condition of understanding: unavoidable though problematic, like interpretation itself. A statement about the forces at work in the present always implies a future, and method requires only that we proceed with some awareness of the tendentiousness of our procedure. But if our subject is not the progress of a national economy or the fortunes of a political party, if our subject is method itself—method as manifested in some of the varieties of contemporary scholarship—then the observer's difficulties are compounded. He may labor to imagine a future toward which present scholarly activities will lead, as cause leads to effect. But more important than the attempt to decide which causal sequences, among the many imaginable, correctly define the present is another question: What kind of causation is at work in the movement of ideas? Our understanding of the intellectual activity of earlier centuries does not usually take the form of causal reconstruction; characteristically, our discussion of an age focuses on a few commanding projects which both sum up and transcend the activities of many predecessors and contemporaries. Imagining causal sequences has not seemed, nor does it now seem, the most appropriate way of undertaking definition and assessment of scholarly activity. One does better to look for symbolic rather than causal relationships, for events which signify the configurations of contemporary scholarship. Such events may seem the very opposite of the hidden causes which clever and determined historians are supposed to seek, since organizers and participants may unabashedly declare the historic importance of

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

their activities; but their very self-consciousness about the symbolic function of the event is itself part of the general state of consciousness that the event claims to record and promote.

Such an event was the first congress of the International Association for Semiotic Studies, held in Milan in 1974. Even if none of the participants learned anything or altered in consequence the nature or direction of his research, the presence of about 650 committed or bemused scholars at a congress of this sort made it an event and testified to a new articulation of scholarly activity. If 650 people attend conferences on semiotics, that does not necessarily cause mutations in the world of scholarship, but it is a fact of symbolic importance. Semiotics, the science of signs, becomes something to be reckoned with, even for those who reject it as a Gallic or a technological obfuscation. And of course when a discipline establishes an organization with committees, officers, publications, when it distributes titles and responsibilities to its adepts, it imposes itself on the scholarly world in symbolic fashion. The proliferation of offices and committees probably inhibits scholarship more than it promotes it, but it does give a discipline an effective presence in the symbolic system of academic research.

The establishment of a new discipline within the system of academic research is not a frequent event. Generally new arrivals explicitly identify themselves as subdivisions of old disciplines and simply undertake to organize more rationally and to pursue more vigorously an existing line of research. The emergence of a discipline like semiotics, however, cannot be guaranteed to leave other disciplines unaffected. Not only would they lose momentum if those who previously called themselves linguists, anthropologists, sociologists, literary critics, philosophers, and so forth, were to identify themselves as semioticians, but the nature of these other disciplines would alter as they lost various kinds of specificity. What might previously have been secondary features of a discipline, what might have seemed simple consequences of its interest in particular objects, might become defining characteristics as semiotics offers another approach to phenomena whose domain previously sufficed to identify a species of scholarly research. Disciplines of *les sciences humaines*—to avoid the English phrasing with its futile attempt to distinguish the humanities from the social sciences—are not autonomous activities but elements of a system with gaps, redundancies, special relationships and indeterminacies. The emergence of a new and aggressive discipline involves a complex readjustment of boundaries and points of focus; no discipline can assume immunity from the effects of this process.

A discussion of the nature and role of semiotics, whose emergence is, at various levels, an event in the world of contemporary scholarship, cannot therefore be simply an account of the methods and content of this particular discipline. If one reflects on the significance of semiotics one must consider the way in which research and writing in the humanities and social sciences are affected by the presence of a new articulation of knowledge: new objects, questions, or criteria. First of all, as a discipline makes a place for itself it makes a past for itself, claiming certain scholars as precursors, interpreting their work in a new light, identifying and redefining forces previously at work in older disciplines and now come into their own as semiotics. To proclaim the advent of semiotics as an event in contemporary scholarship is simultaneously to identify those who

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

will be honored as pioneers and to delineate the failures of various disciplines to deal with the problems that semiotics confronts.

Second, a new discipline has broad implications for contemporary scholarship by the questions it asks and the kinds of answer it seeks. Although it does not set standards for other disciplines, by arguing explicitly for its methods and purposes it does bring to the fore criteria and preoccupations which become relevant to the discussion of other disciplines. More specifically, anthropologists, literary critics, linguists, and others are affected in that they must decide whether to move toward a semiotic viewpoint, to oppose it, or to argue that they have always been doing what now masquerades under a new name.

Finally, a new discipline projects a future. By announcing ambitious programs, which it can do with a more sincere fervor than long-established disciplines which have had a chance to fulfill their promises, semiotics calls upon other disciplines to justify themselves by offering their own visions of the tasks to be accomplished and compels them, in some measure, to take a position on the issues to which it proposes to dedicate itself. Offering a program, it leads opponents or sceptics to say whether its goals are worthwhile and to argue about the scope and validity of the methods with which it proposes to attack these problems. And so in centering my discussion on the pursuit of signs I am engaged less in an explicit evaluation of semiotics as a discipline than in a consideration of the way that reflection on signs is affecting contemporary scholarship in the humanities and social sciences. The emergence of a new discipline, as I have said, creates a past, articulates a present, and projects a future, but to discuss these three activities in turn, as I shall do, is not to attempt anything like a history of semiology: in each case, whether the ostensible topic be the past, present, or future, the real subject is the implications of thinking about signs, the prospects and difficulties which this perspective discloses.

Reflection on signs and meaning is, of course, nothing new. Philosophers and students of language have of necessity always discussed signs in one way and another, and the advent of semiotics has helped to reveal, for example, that what had previously been sneered at as medieval scholasticism was in many respects a subtle and highly developed theory of signs. But, with this exception, until recently the discussion of signs was always ancillary to some other enterprise, usually a discussion of language or of psychology. There had been no attempt to bring together the whole range of phenomena, linguistic and nonlinguistic, which could be considered as signs—no attempt to make the problem of the sign and its varieties the center of intellectual enquiry. Now that people are attempting to do this and have given the name of *semiotics* or *semiology* to their pursuit, one effect is to cast into prominence, as predecessors to be honored, two men who in the early years of the century, envisaged a comprehensive science of signs: the American philosopher Charles Sanders Peirce and the Swiss linguist Ferdinand de Saussure.

Prior to the development of semiology Peirce was coupled with William James as one of the founders of American Pragmatism, whose doctrine maintained, among other things, that the meaning of a conception lay in its practical consequences or implications, and that practical effects determined the validity of concepts. Now, however, it is Peirce's recognition that mental activity of all

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

kinds involves the production and interrelation of signs and that semeiotic, as he called it, would be the science of sciences, which become his major contribution. But Peirce's reflections on the elements of sign situations and the different varieties of each element grew more complex with each reconsideration. There were, he decided, ten trichotomies by which signs can be distinguished, but since there were dependencies and correlations among them, "instead of making 59,049 classes, there will only come to 66." Sixty-six was still too many for his successors. The breadth and complexity of Peirce's scheme, not to speak of the elaborate terminology invented to characterize his 66 types of sign, have discouraged others from entering his system and exploring his insights. His work exerted little influence, and it is only retrospectively that Peirce becomes a major figure.

The other ancestor of semiotics, Ferdinand de Saussure, is a wholly different case. Saussure devoted very few remarks to semiology, as he called it, but in his *Course in General Linguistics*, which laid the foundation for the study of language in the twentieth century, he insisted that language should be studied as a system of signs and that therefore linguistics was only part of a general science of signs. We can imagine, Saussure declared, "a science which would study the life of signs within society. . . . We call it *semiology* from the Greek *semeion* ("sign"). It would teach us what signs consist of, what laws govern them. Since it does not yet exist we cannot say what it will be, but it has a right to existence; its place is insured in advance."[1]

Instead of Peirce's desperate attempts to make all the necessary discriminations among signs, Saussure limited himself to outlining a perspective, noting that "rituals, customs, etc., will, we believe, appear in a new light if they are studied as signs, and one will come to see that they should be included in the domain of semiology and explained by its laws." His influence, therefore, was due not to his own concrete labors in the field of semiology but to the fact that the linguistics which he fathered was eminently suited to serve, eventually, as a model for the semiology of which it is in principle a part. Saussure produced a map for linguistics, articulated its major distinctions, and defined its tasks in ways which came almost to be taken for granted as inherent in the very nature of the subject. In any event, linguistics prospered, and the Saussurian model implicit in its procedures came to seem a model of scientific procedure in the humanities, a model which other disciplines might try to emulate. The structural linguistics which Saussurian thinking had fathered stimulated structuralist movements in other fields, and once structuralism became a significant intellectual movement whose theories and methods were a subject of debate, it became clear that much of the work it embraced was a verison of what Saussure had called semiology: a fulfillment of the prophecies of Peirce and Saussure. They became, belatedly, fathers by adoption—adopted by children seeking legitimation.

The creation or adoption of fathers is a traditional intellectual activity, and Saussure and Peirce were certainly worthy choices; but one suspects that semiology could well have defined itself without them, as the logical outcome of an intellectual reorientation which had been under way for some time. In 1945 the philosopher Ernst Cassirer wrote that "in the whole history of science there is

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

perhaps no more fascinating chapter than the rise of the new science of linguistics. In its importance it may very well be compared to the new science of Galileo which in the seventeenth century changed our whole concept of the physical world."[2] For Cassirer what was revolutionary in linguistics was the primacy granted to relations and systems of relations. Noises that we make have no significance by themselves; they become elements of a language only by virtue of the systematic differences among them, and these elements signify only through their relations with one another in the complex symbolic system we call a language. But if linguistics simply told us this about language it could scarcely have the impact which Cassirer's hyperbolic comparison claimed for it.

To be comparable to Galileo's new science, linguistics would have to change the way in which we think of the universe, or at least of the social and cultural universe. For this to happen, it would have to become a model for thinking about social and cultural activities in general. In short, now that semiology exists it is easy to see that Cassirer's statement implicitly predicts what semiotics explicitly does: that we come to think of our social and cultural world as a series of sign systems, comparable to languages. What we live among and relate to are not physical objects and events; they are objects and events with meaning: not just complicated wooden constructions but chairs and tables; not just physical gestures but acts of courtesy or hostility. If we are to understand our social and cultural world we must think not of independent objects but of symbolic structures, systems of relations which, by enabling objects and actions to have meaning, create a human universe. Several major works of the period between the two World Wars—Cassirer's *The Philosophy of Symbolic Forms*,[3] Alfred North Whitehead's *Symbolism: Its Meaning and Effects*,[4] and Susanne K. Langer's *Philosophy in a New Key*[5]—forcefully asserted the primacy of the symbolic dimension in human experience. Today the configuration of scholarship allows us to see that semiotics, which seeks to describe the underlying systems of distinctions and conventions that enable objects and activities to have meaning, is the systematic fulfillment of a reorientation which they began to describe.

But we can also argue, organizing the history of our modernity from the perspective of the sign, that the crucial insights which semiotics develops lie further back, in the work of Marx, Durkheim, and Freud, who insisted on the primacy of social facts. Human reality cannot be described as a set of physical events, and in focusing of social facts, which are always of a symbolic order, Marx, Freud, and Durkheim dramatically showed that individual experience is made possible by the symbolic systems of collectivities, whether these systems be social ideologies, languages, or structures of the unconscious.

The important question here—as I have said—is not whether one can construct a causal chain of true precursors, or which authors and works should be included in the genealogy of semiotics. The point is that semiotics enables us to perceive in recent intellectual activity a general tendency, variously stated and of differing degrees of explicitness, to stress the role of symbolic systems in human experience and thus to think in terms not of autonomous objects but of systems of relations. Semiotics, in its own historical perspective, becomes the attempt to exploit these insights systematically by identifying and investigating a variety of sign systems; but it is only the insights generated in the emergence of semiotics in recent years that enable one to discover these insights in predecessors.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

The case of Claude Lévi-Strauss, who more than anyone else is responsible for the development of structuralism in fields outside linguistics, illustrates nicely both the major principles of structuralist-semiotic thought and the curious ways in which such insights may emerge. Lévi-Strauss began not with a reading of Saussure, an acquaintence with linguistics, or a desire to investigate the symbolic codes of a society; he began hiking over hills and exploring their geological configurations. His eloquent chapter on "The Making of an Anthropologist" in *Tristes Tropiques* cites as the paradigm of the intellectual quest the moment when, to the geological eye, apparent chaos becomes intelligible; "space and time become one. . . . I feel myself to be steeped in a more dense intelligibility, within which centuries and distances answer each other and speak with one and the same voice." In interpreting terrain the geologist may imagine a history, but "unlike the history of the historians, that of the geologist is similar to that of the psychoanalyst, in that it tries to project in time—rather in the manner of a *tableau vivant*—certain basic characteristics of the physical or mental universe." To geology and Freud, Lévi-Strauss adds a third master, Marx (who "established that social science is no more founded on the basis of events than physics is founded on sense data"). "All three demonstrate that understanding consists in reducing one type of reality to another; that the true reality is never the most obvious; and that the nature of truth is already indicated by the care it takes to remain elusive."[6]

One would expect someone with an interest in geological signs to take historical reconstruction as his goal and to assume that the sign relation was one of effect to cause. Someone interested in psychoanalysis might investigate symptoms as signs of prior causes and seek to reconstruct a history of traumatic events. Or again, a Marxist might see his task as interpreting social phenomena as signs of the events of economic history which caused them. In all three cases, that is to say, there is a temptation to think of interpretation as involving historical reconstruction. Lévi-Strauss's initial insight, without which much of the scholarship of our time might have been different, was the realization that the three cases were not versions of a single reality called history and based on causation. Each discipline employs a temporal projection of a very different sort to describe what is essentially a structure in a system. The eons of the geologist translate the interrelation of strata he sees before him. The economic trends of a century are for the Marxist a projection of the contradictions he finds in a social system. And what the psychoanalyst projects as events of an infantile period may be purely psychic occurrences of very vague temporality which form an unconscious structure. Lévi-Strauss saw that the various temporal projections concealed a common act of understanding which reduced phenomena to structures of a system, to a model or *tableau vivant* as he calls it. From the unlikely starting point of geology he had developed what was to be a basic principle of structuralist analysis: to understand phenomena is to reconstruct the system of which they are manifestations.

This insight is consolidated in a famous article of 1945 entitled "Structural Analysis in Linguistics and Anthropology" in which Lévi-Strauss argues that anthropologists might learn more directly from linguistics what he had discovered in more oblique fashion. Linguistics, and particularly phonology, which was the most striking early success of structural linguistics, "ought to play the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

same renovating role for the social sciences that nuclear physics, for example, played for the exact sciences." In reducing the apparent chaos of speech sounds to an order, phonology moved "from the study of conscious linguistic phenomena to that of their unconscious infrastructure."[7] A speaker of a language is not consciously aware of the phonological system of his language, but an unconscious system of distinctions and oppositions must be postulated to account for the fact that he interprets two physically different sound sequences as instances of the same word, yet in other cases distinguishes among sequences which are acoustically very similar. Phonology reconstructs an underlying system, and in so doing it focuses not on terms or individual elements but on relations. Sounds of a language are not defined by some essential properties but by a series of functional distinctions. I can pronounce *cat* in various ways so long as I maintain a distinction between *cat* and *bat*, *cut*, *cad*, etc. The example of linguistics, Lévi-Strauss argues, teaches the anthropologist that he should try to understand phenomena by considering them as manifestations of an underlying system of relations. To describe that system would be to identify the oppositions which combine to differentiate the phenomena in question.

Lévi-Strauss's studies of marriage rules in *The Elementary Structures of Kinship*,[8] his work on *Totemism*, his discussion of various logics in *The Savage Mind*, and his monumental four-volume study of North and South American mythology, all follow, in various ways, this procedure. The marriage practices of various societies are reduced to systems of rules, and these systems are themselves described as variant realizations of a limited set of elementary oppositions. *The Savage Mind* and *Totemism* argue that anthropologists have often failed to understand the thought and behavior of their subjects because they have attempted atomistic and functionalist explanations, taking phenomena one by one rather than treating them as part of an underlying system with a logic of its own. If a particular clan has the bear as its totem, one need not indulge in far-fetched religious, historical, or economic explanations. "To say that clan A is 'descended' from the bear and clan B is 'descended' from the eagle is only a concrete and abbreviated way of stating the relationship between A and B as analogous to the relationship between the two species."[9] Bear and eagle are logical operators, concrete signs, and to understand them is to analyze their place in a system of signs.

The study of myth is Lévi-Strauss's most ambitious project, because the myths he investigates appear totally bizarre and inexplicable, full of the most unaccountable incidents, characters, metamorphoses. As he explains in the preface to his first volume, *The Raw and the Cooked*, "if it were possible to show that the apparent arbitrariness of myths, the supposed freedom of inspiration, the seemingly uncontrolled process of invention, implied the existence of laws operating at a deeper level, then the conclusion would be inescapable . . . if the human mind is determined even in its creation of myths, *a fortiori* it is determined in other spheres as well."[10] The chaos of myths is a challenge to the mind nurtured on geology, Marx, and Freud, and the quest for an underlying system which would order their multifarious forms involves treating them as manifestations of a "language" whose fundamental units and oppositions he must identify.

In isolating fundamental oppositions, such as raw–cooked, day–night, sun–moon, and many of more exotic and unexpected sorts Lévi-Strauss is describing

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

codes: sets of categories drawn from a single area of experience and related to one another in ways that make them useful logical tools for expressing other relations. His method shows that within the most bizarre incidents can be found categories which, because of their relations to other categories within codes, have an expressive function. The general implication of this method, which has become a fundamental principle of structural and semiotic analysis, is that elements of a text do not have intrinsic meaning as autonomous entities but derive their significance from oppositions which are in turn related to other oppositions in a process of theoretically infinite semiosis. If a text compares a woman to the moon, that predication has no inherent meaning; significance depends on the opposition between sun and moon, either or both of which may have other correlations within the text itself, within other related texts, and within the general symbolic codes of a culture. The relational nature of signs produces a potentially infinite process of signification.

There is still much debate about precisely what Lévi-Strauss has achieved in his analysis of myths. Anyone who reads his discussions of these incomprehensible tales can see that he has discovered an underlying logic, though it is not clear whether one could ever in principle or in practice show that it is *the* logic of myths. What is crucial for contemporary scholarship, however, is another methodological issue which his work has raised and which can reorganize one's conception of research in *les sciences humaines*.

In speaking of understanding as a process of reducing one type of reality to another, Lévi-Strauss explicitly avoids the model of causal explanation. The type of explanation which he offers in his analysis may on occasion be projected in time and treated as a causal analysis, but that is never its central or defining feature. Structural explanation, as it seems best to call it, relates objects or actions to an underlying system of categories and distinctions which make them what they are. In this perspective, to explain phenomena is not to discover temporal antecedents and link them in a causal chain but to specify the place and function of the phenomena in a system.

This is one of the more important lessons that linguistics has offered other disciplines of the humanities and social sciences: that without opposing the notion of causal explanation, they can shift from a historical to an ahistorical perspective and attempt to describe systems rather than trace the antecedents of individual events. The distinction between what linguistics calls synchronic description—the analysis of a system without respect to time—and diachronic analysis—the attempt to construct a historical evolution—has become a major criterion in characterizing research. Increasingly, even in fields where scholarship was previously presumed to mean historical research, attention has turned to synchronic analyses. To understand social and cultural phenomena, whether they be congressional committees, neckties, or cross-country skiing, is not to trace their historical evolution but to grasp their place and function in various systems of activity and to identify the distinctions which give them significance. The single most important way of characterizing a piece of research may be to ask whether it is synchronic or diachronic in character, and it is the structuralist perspective which has given this question such prominence.

Lévi-Strauss's frequent references to linguistics, and his announcement in

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

his inaugural lecture at the Collège de France in 1961 that he saw anthropology as part of semiology, cast him in the role of a scholar engaged in the pursuit of signs, but the role and status of signs in his work is a problem of some complexity. Lévi-Strauss always deals with symbolic systems and attempts to reconstruct the codes which constitute these systems: but whereas a linguist analyzing a language asks what are the rules and conventions which enable sequences of sounds to have the meanings they do for members of a culture, Lévi-Strauss does not ask what are the codes that account for the meanings that myths have within a particular culture. Though he does often draw upon ethnographic information in isolating functional oppositions, he is constructing a transcultural logic of myth. He is not interested in elements of myths as signs within a particular native culture. Myths are always and primarily signs of the logic of myth itself. "I do not aim," he says, "to show how men think in myths, but how myths think in men, unbeknownst to them."[11]

Lévi-Strauss's paradoxical formulation poses a question which, as we shall see below, is central to the semiotic perspective: what is it that enables one to say that language speaks, myth thinks, signs signify? But in neglecting the opportunity to study myths as signs within a particular culture, Lévi-Strauss is bypassing the central activity of semiotics as it has recently taken shape. Treating as signs objects or actions which have meaning within a culture, semiotics attempts to identify the rules and conventions which, consciously or unconsciously assimilated by members of that culture, make possible the meanings which the phenomena have. Information about meaning—whether particular actions are considered polite or impolite, whether a musical sequence seems resolved or unresolved, whether an object connotes luxury or penury—is therefore crucial since what the analyst wishes to do is to isolate those distinctions which are responsible for differences of meaning.

For example, a semiologist, sociologist, or ethnologist (the way that semiotics displaces traditional disciplinary boundaries becomes obvious here) who set out to study clothing in a culture would ignore many features of garments which were of great importance to the wearer but which did not carry social significance. To wear bright garments rather than dark may be a meaningful gesture, but to opt for brown rather than gray might not. Length of skirts might be a matter of purely personal preference whereas choice of materials would be rigidly codified. In attempting to reconstruct the system of distinctions and rules of combination which members of a culture display in choosing their own garments and in interpreting those of others as indications of a particular lifestyle, social role, or attitude, the semiotician would be identifying the distinctions by which garments become signs.

Whatever area he is working in, someone adopting the semiotic perspective attempts to make explicit the implicit knowledge which enables people within a given society to understand one another's behavior. Often, of course, this implicit knowledge is a deeply rooted set of cultural norms and conventions which operate subconsciously and which members of a culture might angrily deny. In these cases, the description of a semiotic system becomes an act of demystification, of exposure. The pleasure of revealing the culturally determined nature of behavior has doubtless been the impetus behind much semiotic analysis, but one would be mystified by the demystification itself if one thought that the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

description of semiotic systems made the individual more free or that the semiotic analysis was in any way inspired by the prospect of liberating man. On the contrary, structuralist and semiotic thinking has been repeatedly labeled "anti-humanistic," and Michel Foucault has provided a target for such attacks in maintaining that "man is only a recent invention, a figure not yet two centuries old, a simple fold in our knowledge" which will soon disappear.[12]

What does the pursuit of signs have to do with the disappearance of man? A whole tradition of thought treats man as essentially a thinking being, a conscious subject who endows objects around him with meaning. Indeed, we often think of the meaning of an expression as what the subject or speaker "has in mind." But as meaning is explained in terms of systems of signs—systems which the subject does not control—the subject is deprived of his role as source of meaning. I know a language, certainly, but since I need a linguist to tell me what it is that I know, the status and the nature of the "I" which knows is called into question. "The goal of the human sciences," says Lévi-Strauss, "is not to constitute man but to dissolve him."[13] Although they begin by making man an object of knowledge, these disciplines find, as their work advances, that the self is dissolved as its various functions are ascribed to impersonal systems which operate through it.

"The researches of psychoanalysis, of linguistics, of anthropology have 'decentered' the subject in relation to the laws of its desire, the forms of its language, the rules of its actions, or the play of its mythical and imaginative discourse," writes Foucault.[14] The distinction between man and the world is a variable one that depends on the configurations of knowledge at a given time, and the various disciplines engaged in semiotic analysis have treated as systems of conventions so much of what used to be the property of the thinking subject that any notion of man founded thereon becomes problematic. As the self is broken down into component systems, deprived of its status as source and master of meaning, it comes to seem more and more like a construct: a result of systems of convention. Even the idea of personal identity emerges through the discourse of a culture: the "I" is not something given but comes to exist as that which is addressed by and relates to others. In short, as Jean-Marie Benoist puts it in his account of *La Révolution structurale*, what theoretical investigation discovered was not man but signs. No longer man-centered, "contemporary research could conceive of itself as a semiotics: semiotics of the unconscious for Lacan, semiotics of the codes of kinship and myth for Lévi-Strauss, semiotics of the relations and contradictions in society for Althusser, semiotics of literature for Barthes and Genette, and a semiotics of historical discourse and documents, of which Foucault provides the discourse on method."[15] In each case, explanation involves the description of systems of signs, not the tracing of an event to a mind which would be allowed to count as its source.

If the "dissolution" of man into a series of systems is the ultimate effect of a structuralist and semiotic perspective, if everything which is meaningful within human cultures can be treated as a sign, then as Benoist's list already suggests, semiotics embraces a vast domain: it moves in, imperialistically, on the territory of most disciplines of the humanities and social sciences. Any sphere of human activity, from music to cooking to politics, can be an object of semiotic study;

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

and it is precisely because any signifying activity calls for semiotic investigation that the emergence of semiotics may bring about a major reorganization of scholarly research. If the study of music as a sign system is assimilated to semiotics, the other aspects of musicology will form a discipline which must define itself in a new way.

Whatever configurations the future holds, semiotics at the moment brings together a whole series of projects which study signs but which it does not know how to classify. Umberto Eco's recent *A Theory of Semiotics* offers a list of current concerns which is amusing in its very disorder: Zoosemiotics, Olfactory signs, Tactile communication, Codes of taste, Paralinguistics, Medical semiotics, Kinesics and proxemics, Musical codes, Formalized languages, Written language, Unknown alphabets and secret codes, Natural languages, Visual communication, Systems of objects, Plot structure, Text theory, Cultural codes, Aesthetic texts, Mass communication, Rhetoric.[16]

As this list makes abundantly clear, one of the major tasks that semiotics must face is to organize itself. This is in fact its primary concern, since it involves determining what are the principal varieties of sign and how they relate to one another. But this is a question which will be confronted by other disciplines as well as they react to the imperialistic claims of semiotics. If they welcome assimilation to or association with semiotics they will try to determine how signification takes place in the sphere of human activity that concerns them and how these processes relate to others studied by semiotics; but if they resist the claims of semiotics they will also engage in debate about the processes of signification that concern them. The result of these various efforts will doubtless be a demonstration of the inadequacy of the categories and distinctions which semiotics has so far proposed for classifying signs. As Julia Kristeva has written, "semiotics cannot develop except as a critique of semiotics. At every moment in its development semiotics must theorize its object, its own method, and the relationship between them; it therefore theorizes itself and becomes, by thus turning back on itself, the theory of its own scientific practice. . . . It is a direction for research, always open, a theoretical enterprise which turns back upon itself, a perpetual self-criticism."[17]

The best way to illustrate the complex and self-reflexive progress of a semiotic enterprise is to consider what semiotics has done and promised to do for the study of the most complex of sign systems, literature. Literature is the most interesting case of semiosis for a variety of reasons. Though it is clearly a form of communication, it is cut off from the immediate pragmatic purposes which simplify other sign situations. The potential complexities of signifying processes work freely in literature. Moreover, the difficulty of saying precisely what is communicated is here accompanied by the fact that signification is indubitably taking place. One cannot argue, as one might when dealing with physical objects or events of various kinds, that the phenomena in question are meaningless. Literature forces one to face the problem of the indeterminacy of meaning which is a central if paradoxical property of semiotic systems. Finally, unlike so many other systems which are devoted to ends external to themselves and their own processes, literature is itself a continual exploration of and reflection upon signification in all its forms: an interpretation of experience; a commentary on

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

the validity of various ways of interpreting experience; an exploration of the creative, revelatory, and deceptive powers of language; a critique of the codes and interpretive processes manifested in our languages and in previous literature. Insofar as literature turns back on itself and examines, parodies, or treats ironically its own signifying procedures, it becomes the most complex account of signification we possess.

Literary criticism, as a result of the semiotic perspective, has devoted much time to showing that this is so. In interpreting a novel like Flaubert's *Madame Bovary*, for example, one would show how irony, which is itself a process of signification, works to undermine other types of signification: both Emma's reading of her own experience and the reader's attempt to make events and descriptive details fit into "novelistic" signifying patterns. One can show that the novel, like all discourse which takes the initial step of calling itself fictive and thereby poses inescapably the problem of its signifying status, is ultimately "about" signs and meaning.[18] Or again, in interpreting a poem like Blake's *London*, one would argue that though in one sense it is a vision of urban misery, it narrates, at a literal level, acts of interpretation, a reading of signs; and the rhetorical figures which the poem uses to narrate the vision ("the hapless soldier's sigh/Runs in blood down palace walls" and "the youthful harlot's curse/ . . . blights with plagues the marriage hearse") because they are so very unusual, pose the question of the status of the interpretive acts here recounted. If we imagine a speaker walking through London streets reacting in this way to sighs and curses, then we have an account of a signifying process gone wrong, an obsessional imagination; in order to accept and make sense of the interpretations which the poem offers we must treat as fictions the interpretive acts which the poem describes. In leading the reader to choose an interpretive strategy, the poem explores the paradoxical qualities of fictional discourse and the undecidable nature of figurative language.

Other literary works, of course, are much more violently explicit in their dealings with signs and signification, transgressing in their own several ways all the linguistic and discursive conventions one can think of. As explorations both of the power of language to create thought and of the limits of discourse, works of this sort constitute a radical contribution to a theory of signs and signification, for they show the impossibility of treating signification as a purely code-like phenomenon. When they appear in literature, as they do, new lexical items will be given some kind of meaning by readers (consider Joyce's "Stay us wherefore in our search for tighteousness" or Carroll's "brillig") and syntactic combinations one would have thought impossible will be interpreted (Cummings's "Anyone lived in a pretty how town/ with up so floating many bells down"). Criticism attuned to semiotics interprets works as semiotic explorations.

But if literary works make it clear that one cannot set limits to the signifying process and define once and for all the appropriate system of conventions, they also provide conclusive evidence for the existence of a semiotic system which makes literature possible. Just as violations of etiquette testify to the existence of conventions which make it possible to be polite or impolite, so the flaunting of linguistic and literary conventions by which literary works bring about a renewal of perception testifies to importance of a system of conventions as the basis of literary signification. Precisely because literary works so often attempt to

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

achieve their effects by parodying or treating ironically previous conventions, by ending in unexpected ways or using details in ways which will surprise readers, it is important, for any account of literature, to try to define the underlying systems of convention, to characterize literature as an institution.

One must distinguish between the kind of interpretive criticism discussed earlier, which interprets individual works as statements about signification, and this semiotics of literature, which does not interpret works but tries to discover the conventions which make meaning possible. Here the goal is to develop a poetics which would stand to literature as linguistics stands to language. Just as the task of linguists is not to tell us what individual sentences mean but to explain according to what rules their elements combine and contrast to produce the meanings sentences have for speakers of a language, so the semiotician attempts to discover the nature of the codes which make literary communication possible.

This project, whether explicitly defined as it was by Tzvetan Todorov, or elegantly and obliquely pursued as it was by Roland Barthes and Gérard Genette, inspired a great deal of highly original critical and theoretical discussion of literature in France in the 1960s.[19] Precisely why this should have happened in France is difficult to say. Anglo-American thinking, especially literary thinking, had long assumed a radical distinction between the arts and the sciences, which made critics reluctant to ape the procedures of scientists, offering comprehensive or systematic theories, creating technical terms, and generally seeking the adventure of abstract thought. The French, exposed to Lévi-Strauss's on *La Pensée sauvage* and to the writings of Gaston Bachelard, philosopher of science turned student of literary imagery, reveled in neologisms, proposed systematic frameworks, and generally indulged in invention. "I passed through a euphoria of scientificity," said Roland Barthes.[20] The products of such activity are often rebarbative and tendentious, but a situation of intellectual excitement does produce ideas.

Poetics or the semiotics of literature is no longer exclusively or even primarily a French concern. Work on plot structure, the goal of which is a grammar of plots, has been carried out in many countries. Work on narrative codes or techniques can assimilate to recent French discussions like Genette's *Discours du Récit*[21] research done earlier in a different intellectual context in Germany and the United States. Indeed, today the semiotics of literature is very much an American and European phenomenon organized into loose groups of varying persuasions but not into national schools. It is governed by the assumption that a systematic theory of discourse if not of literature (for one of the effects of semiotics is to question the distinction between literary and nonliterary discourse) is possible, though there may be little agreement about precisely what "languages" (information theory, semantics, systems theory, psychoanalysis) will be most useful in establishing the categories and identifying the codes of the discursive systems at work in texts. A major point on which there would be agreement, however, is that literary works are to be considered not as autonomous entities, "organic wholes" as American New Criticism put it for so long, but as intertextual constructs: sequences which have meaning in relation to other texts which they take up, cite, parody, refute, or generally transform. A text can be read only in relation to other texts, and it is made possible by the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

codes which animate the discursive space of a culture. Literary study, therefore, no longer concentrates on authors; the work is a product not of a biographically defined individual about whom information could be accumulated, but of writing itself. After all, to write a poem the author had to take on the character of poet, and it is that semiotic function of poet or writer rather than the biographical function of author which is relevant to discussion of the text.

Literary study experienced what Barthes called "the death of the author" but almost simultaneously it discovered the reader, for in an account of the semiotics of literature someone like the reader is needed to serve as center. The reader becomes the name of the place where the various codes can be located: a virtual site. Semiotics attempts to make explicit the implicit knowledge which enables signs to have meaning, so it needs the reader not as a person but as a function: the repository of the codes which account for the intelligibility of the text.[22] Because literary works do have meaning for readers, semiotics undertakes to describe the systems of convention responsible for those meanings.

This is a coherent and necessary program: since communication does take place we must discover how it occurs if we wish to understand ourselves as social and cultural beings. But literature itself, in its continual pressure on and violation of codes, reveals a paradox inherent in the semiotic project and in the philosophic orientation of which it is the culmination. To account for the signification of, shall we say, a metaphor is to show how the relationship between its form and its meaning is already virtually present in the systems of language and rhetoric. The metaphor itself becomes not a radical or inaugural act but a manifestation of a preexistent connection. Yet the value of the metaphor, the value of our experience of the metaphor, lies in its innovatory, inaugural force. Indeed, our whole notion of literature makes it not a transcription of preexisting thoughts but a series of radical and inaugural acts: acts of imposition which create meaning. The very conventions to which we appeal in explaining literary meanings are products: products which, it would seem, must have acts as their source.

This second perspective deconstructs the first; it seems to bring about a reversal, explaining meaning not by prior conventions but by acts of imposition. However, the first perspective also deconstructs the second in its turn, for acts of imposition are themselves made possible by the situations in which they occur, and meanings cannot be imposed unless they are understood, unless the conventions which make possible understanding are already in place. The semiotics of literature thus gives rise to a "deconstructive movement" in which each pole of an opposition can be used to show that the other is in error but in which the undecidable dialectic gives rise to no synthesis because the antinomy is inherent in the very structure of our language, in the possibilities of our conceptual framework.

What happens in literary semiotics is but one version of a general situation which is gradually coming to be recognized as an inescapable feature of our ways of thinking about texts and signification. Semiotics is the instrument of this revelation because it is the logical culmination of what Jacques Derrida calls the "logocentrism" of Western culture: the rationality which treats meanings as concepts or logical representations that it is the function of signs to express. We speak, for example, of various ways of saying "the same thing."

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Semiotics begins as a critique of the logocentric assumption that concepts exist prior to and independently of their expression. In analyzing signification Saussure and his later followers insist that forms and concepts do not exist independently of one another, but that the sign consists of the union of a signifier and signified. Moreover—and this is the important point—both signifiers and signifieds are purely relational entities, products of a system of differences. To speak of the concept of "brown," for example, is, according to semiotics, a way of referring to a complex network of oppositions which articulates the spectrum of colors on the one hand and the spectrum of sound on the other. The meaning of *brown* is not a representation in my mind at the moment of utterance but a space in a complex network of differences.[23]

Semiotics thus takes up the problem of the sign, on which logocentric notions of signification have been based, and gives it a relational or differential interpretation which seems to not only make possible a new type of explanation—structural explanation in terms of underlying systems of relation—but also to displace logicentrism. However, as Derrida has shown, in a series of works which are the most brilliant products of a semiotics whose limits they describe and attempt to transcend, semiotics does not escape logocentrism: though the source of meanings is no longer a consciousness in which they exist prior to their expression, their source becomes a system of differences which semiotics treats as the necessary condition of any act of signification.[24] This is a marked advance, a far more adequate account of signification than that which it originally set out to criticize, but it ultimately encounters the same formal difficulty: instead of depending on the prior existence of a system of concepts, expression now depends on the prior existence of a system of signs.

One deconstructs this perspective by arguing that the differences ultimately responsible for meaning did not simply fall from heaven but are themselves products. Acts of signification are necessary to create signifying differences. But this perspective gives rise to no discipline; it is not a position that can be maintained because if one tries to discuss acts of signification one immediately is led to describe the oppositions which enable an act to signify; one inevitably finds oneself back in the semiotic perspective, describing a system. This irresolvable dialectic, the solidarity of these contradictory perspectives, Derrida captures with the term *differance*, which is both a difference and a differing, designating a passive difference always already in place as the ground of signification and an act of differing which produces the differences it presupposes. To spell *differance* with an *a* instead of an *e* is of course to press against the limits of a logocentric language, but the concept it produces can be understood only in our language, in which, of course, it is a contradiction. More generally, escape from logocentrism is impossible because the language we use to criticize or to formulate alternatives works according to the principles being contested.

The paradoxes and undecidables which this perspective discloses as the unavoidable bases of language and thought are more familiar and more easily exemplified in the realm of literature than elsewhere. The very notion of rhetorical figures, which have now become a major object of critical attention, captures a fundamental paradox.[25] A rhetorical figure is a situation in which language means something other than what it says, a violation of the code. But lest that violation introduce a radical undecidability to linguistic situations, leading us to wonder how we could ever know whether language means what it

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

appears to be saying, these violations are codified, as a repertoire of highly artificial and conventional devices which writers can draw on to produce meaning. What looks at first like an inaugural creative act, a violation of the code, is accounted for by the formulation of a code on which its meaning is said to depend. Nor is this last move an unfortunate error which might have been avoided. The very notion of rhetorical effects—the possibility of metaphorical signification, for example—requires there to be a distinction between literal meaning and metaphorical meaning and hence the beginnings of a rhetorical code.

Thus, even the study of reputedly deviant forms of signification leads one back to the same problematic. Undertaking a rigorous investigation of signs and signification, semiotics produces a discipline which, ultimately, reveals the fundamental contradictions of the signifying process as we understand it. Semiotics leads, necessarily, to a critique of semiotics, to a perspective which shows the errors of its ways. But that perspective is never a viable alternative. It is not a position from which one could undertake an alternative analysis of signs and systems of signs, for the notion of analysis, of explanation, of production of models are all part of the semiotic perspective, and to undertake any of them is immediately to revert to that perspective. The alternative, then, is not a discipline, not another mode of analysis, but acts of writing, acts of displacement, play which violates language and rationality. Though these acts can themselves be analyzed and understood, discussed in terms of codes which make them meaningful, they are in their moment, as examples of the play of signifiers, challenges to perspective whose limitations they expose.

Given this structure, one can think of semiotics and the future it projects in two ways. First, it offers a discipline which can bring together in a comparative perspective a whole range of phenomena which do respond to treatment by a common method. Proposing structural explanation in place of historical and causal reconstruction, making explicit the interdependence of social phenomena on one another by analyzing them in terms of systems of relations, and demonstrating the extent to which what we call man is the juncture of a series of interpersonal systems which operate through him, semiotics offers both methodological claims which will be debated even by those who reject them, and a program which in its ambitious scope will provide a place for any scholar who needs one. Its combination of novelty and comprehensiveness may help it to attract funds, especially since funds in the humanities are often easier to secure when one's project appears part of a cooperative and "scientific" effort. And the fact that the pursuit of semiotics leads to an awareness of its limits, to an awareness of the impossibility of producing a coherent and comprehensive theory of signification, may be an advantage rather than a disadvantage, since to many scholars a sense of limits may be comforting.

But on the other hand, insofar as it leads to the limits of its own theory, semiotics gives rise to a kind of interpretive activity, a Derridean *double science*, a mode of deconstructive reading. Though dependent upon the more sober and difficult attempt to describe systems of signs, it enjoys citing the ultimate impossibility of that activity before turning to the task it has set itself: that of rereading the major texts of Western literature and philosophy as sites on the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

boundaries of logocentrism and showing, in the most subtle interpretations which scholarship has yet produced, how these texts are already riven by the contradictions and indeterminacies which now seem inherent in the exercise of language.[26] That this should be a major project of humanistic scholarship in the coming years is not, perhaps, a fact one should celebrate, but for those who are and will be engaged in it, the activity is one of ironic celebration.

REFERENCES

[1]Ferdinand de Saussure, *Cours de linguistique générale*, 3rd ed. (Paris, 1967), p. 33. All translations are my own unless otherwise indicated.

[2]Ernst Cassirer, "Structuralism in Modern Linguistics," *Word* 1 (1945): 99.

[3]Ernst Cassirer, *The Philosophy of Symbolic Forms* (New Haven, 1953, 3 vols.). Original German publication 1923–1929.

[4]Alfred North Whitehead, *Symbolism: Its Meaning and Effects* (New York, 1927).

[5]Suzanne K. Langer, *Philosophy in a New Key* (Cambridge, Mass., 1942).

[6]Claude Lévi-Strauss, *Tristes Tropiques*, translated by John and Doreen Weightman (Harmondsworth, Middlesex, 1976), pp. 69–70.

[7]Claude Lévi-Straus, *Anthropologie structurale* (Paris, 1958), p. 40. The article originally appeared in *Word*, vol. 1.

[8]Claude Lévi-Strauss, *The Elementary Structures of Kinship*, translated by Rodney Needham et al. (Boston, 1966).

[9]Claude Lévi-Strauss, *Totemism*, translated by Roger Poole (Harmondsworth, Middlesex, 1969), p. 100.

[10]Claude Lévi-Strauss, *Le Cru et le cuit* (Paris, 1966), p. 18. Vol. 1 of *Mythologiques* (Paris, 1964–1967). English translation, *The Raw and the Cooked* (New York, 1969).

[11]Ibid., p. 20.

[12]Michel Foucault, *Les Mots et les choses* (Paris, 1966), p. 15. English translation, *The Order of Things* (London, 1970).

[13]Claude Lévi-Strauss, *La Pensee sauvage* (Paris, 1962), p. 326. English translation, *The Savage Mind* (Chicago, 1966).

[14]Michel Foucault, *L'Archéologie du savoir* (Paris, 1969), p. 22. English translation, *The Archeology of Knowledge* (New York, 1972).

[15]Jean-Marie Benoist, *La Révolution structurale* (Paris, 1975), p. 16.

[16]Umberto Eco, *A Theory of Semiotics* (Bloomington, Ill., 1976), pp. 9-13.

[17]Julia Kristeva, *Semiotike: Recherches pour une sémanalyse* (Paris, 1969), p. 30.

[18]See my *Flaubert: The Uses of Uncertainty* (Ithaca, N.Y., 1974).

[19]For a bibliography and further discussion see my *Structuralist Poetics* (Ithaca, N.Y., 1975) and Tzvetan Todorov, *Poétique* (Paris, 1973). The most important works are Roland Barthes, *S/Z* (Paris, 1970) and the three volumes of Gérard Genette's *Figures* (Paris, 1966, 1969, 1972).

[20]Roland Barthes, "Réponse," *Tel Quel*, 47(1971):97.

[21]Gérard Genette, *Discours du Récit* (vol. 3 of *Figures*, Paris, 1972).

[22]For discussion see Michael Riffaterre, *Essais de stylistique structurale* (Paris, 1971), my *Structuralist Poetics*, and Stanley Fish, "Literature in the Reader: Affective Stylistics," *New Literary History 2* (1970).

[23]See my *Saussure* (London, 1976), Chapter 2.

[24]The clearest exposition of Derrida's position is in his *Positions* (Paris, 1972). Two important works have been translated, *Speech and Phenomena* (Evanston, Ill., 1973) and *Of Grammatology* (Baltimore, 1977), of which the former is more traditionally philosophical. One should also consult the essays collected in *L'Ecriture et la différence* (Paris, 1967) and *Marges de la philosophie* (Paris, 1972).

[25]For discussion see Derrida's "La Mythologie blanche" in *Marges de la philosophie*. An English translation was published in *New Literary History* 6 (1974).

[26]This program for literary criticism is announced, ironically, by Paul de Man in "Semiology and Rhetoric," *Diacritics* 2 (1973) and earnestly by J. Hillis Miller, "Stevens' Rock and Criticism as Cure," II, *Georgia Review* 30 (1976).

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

GEORGE A. MILLER

Problems of Communication

COMMUNICATION is a problem. Sometimes it is even difficult to accomplish, but I mean problem as opposed to discipline; communication has been studied in many ways from various points of view. In the imaginary matrix of problems-by-disciplines, the row devoted to it would have contributions in almost every column.

How did problems of communication become so ubiquitous? Because communication is so ubiquitous, of course. But how did we come to see so many different things as instances of communication? We speak of communicable diseases, communicating arteries, and genetic communication from one generation to the next, as well as animal communication, mass media of communication, interpersonal communication. Some even speak of art as a form of communication. The generality of the concept is reflected in our modern definitions of it: whenever a process at one place or time is correlated with a process at another place or time we can call it communication—even when, as in extrasensory perception, we have no idea what channel might cause a correlation.

Preoccupation with communication has been characteristic of the twentieth century—to locks and clocks, traditional symbols of Western civilization, we have added the mail box. And the telegram, telephone, radio, television, computer, and satellite. Nor has our concern been wholly technological. In academic quarters, the social and behavioral sciences have been almost morbidly interested. Linguistics has emerged as one of the most rigorous fields of study; sociology and political science have developed theories of the mass media and their power to affect public opinion; psychology has probed the processes involved and tried to infer thoughts from the symbols that communicate them; anthropology has explored connections between language and culture. Moreover, neurology has puzzled over the effects of brain lesions on communicative competence, and zoology has speculated about the evolutionary implications of intraspecies and interspecies communication. Even philosophy has contributed theories of the logical and semantic substructures of natural languages.

The various heads of this modern Hydra are not totally independent. They do not communicate with each other as often, perhaps, as one might predict from their common dedication to the processes of communication; but occasionally something develops with implications beyond disciplinary boundaries and then it becomes necessary to get various heads together. Because each head

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

speaks a different jargon, the translation process is sometimes as creative as the original development; much progress has come from combining diverse conceptions from different disciplines into a coherent whole. In an age of analysis, therefore, the study of communication has provided a special refuge for synthesists, for those personalities whose response to disciplinary boundaries is a feeling of intellectual claustrophobia.

Because I have always enjoyed transporting ideas across provincial lines, the two conceptual advances in this general area that I want to discuss here both arose outside my own field of psychology. I choose them not only because both have generated interest and excitement and hold much promise for the future, but because there is tension between them. Because I have devoted a not inconsiderable fraction of my own energies to persuading my fellow psychologists in general, and language psychologists in particular, that our discipline stands to profit from assimilating both of these developments into our own ways of thinking and working, the tensions between them have been particularly acute for me and for those who share my persuasion. Discussing them is, therefore, a form of self-therapy. But I believe the discussion has broader import than that. How the tensions will be resolved, if they can be, could shape our thought and work on psychological aspects of communication for many years to come.

But first I must narrow the discussion to my own areas of competence. I will be concerned here with only the kind of communication that occurs among people by virtue of their use of human language. And the aspect of linguistic communication that is my central interest is psychological: What can the facts of linguistic communication teach us about the nature of the human mind?

At first blush, to speak of language does not seem to narrow our scope much more than did speaking of communication, again because of the generality of modern definitions of language—because we see so many different things as instances of language. Whenever a symbolic code is used to achieve the goals of communication (in the broad sense) we can call it language. Thus we speak of the languages of the bees or the dolphins, of computer languages, and the languages of gesture or music, as well as of the English, Russian, or Chinese languages.

The extended senses of "communication" and "language" reflect a general principle of lexicography that is itself a matter of some psychological interest. Many English nouns have both a literal and a functional meaning. "Table," for example, has the literal sense of piece of furniture with a horizontal top supported by one or more vertical legs, but it also has an extended, functional sense of anything that serves the function of a table. Thus, under appropriate circumstances, one can say that a tree stump or a crate is a good table if it is serving the function of a table, even though, speaking literally, stumps and crates are not tables. People who speak English understand such extensions implicitly, if not explicitly, and can functionally extend a great variety of nominal lexical concepts. Thus, "communication" can easily be understood to denote anything that serves the function of communication, and "language" as anything that serves the function of language. With appropriate qualifications, functional extension can be formulated as a lexicographic rule—and psychologists can argue whether a person who knows English and conforms to the rule can be said to know the rule. But I am getting ahead of my story.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

In most cases, functional extension is an innocent and useful semantic trick, but it can cause trouble if the extended sense is confused with the literal. No one is likely to say, for example, "The merchandise was packed in tables" on the grounds that crates can be called tables by functional extension of "table." But people who say, for example, "Studies of DNA now enable us to understand the language of heredity" are likely to argue rather stubbornly that they are using "language" literally. That is to say, some people fail to distinguish two senses; for them "communication" and "language" have only functional meanings.

In order to narrow my topic, therefore, I will insist on a literal sense of "language": the grammatical strings of spoken words that people use. I would like to complete this sentence more gracefully—"the grammatical strings of spoken words that people use to communicate information," for example—but being explicit about what people use language for raises a host of problems. Communicating information is but one function that human language serves; others are self-expression, entertainment, ritual, record-keeping, reasoning, prayer. A psychology of linguistic communication cannot ignore the many uses to which language is put, but for the moment I want to draw the kind of structure-function distinction we might adopt, say, in discussing walking: what it is versus where it might take you. The distinction seems obvious, yet it can pose problems when we try to sort out which linguistic phenomena are the province of structural theories and which are functional. But again I am getting ahead of my story.

Anthropologists tell us that no human group lacks a language (in this literal sense), and therefore we are led to believe that a proclivity for this form of communication must be a distinctive feature of *Homo sapiens*. This viewpoint sets a problem for psychologists: What is there about the human mind that underlies the human capacity for language? Presumably, some cognitive faculty has gotten out of hand, with results as peculiar and specific to man as long necks are to giraffes.

Descartes has been given considerable credit for noting that speech distinguishes man from other animals—probably because he linked it with the possession of a soul—but the observation is hardly one that would have surprised the ancient Greeks. Today we would not accept the soul as a satisfactory explanation of man's unique linguistic gifts, but providing a scientific substitute is no simple task. Those who try can be crudely divided into two camps. One claims that mankind is characterized by innate, special-purpose, cognitive mechanisms peculiar to language; the other claims that mankind simply has, in greater degree, the same innate, general-purpose cognitive mechanisms that all higher mammals share. This difference is sometimes incorrectly called a disagreement about innateness, but both camps subscribe to innateness; the ability to learn is itself innate. The real disagreement is whether there are special-purpose mechanisms or capacities that play no other role in man's cognitive life.[1]

Attempts to resolve this disagreement have not been particularly enlightening. I personally favor a special-purpose view, but the only way to prove it that I can imagine is to push the general-purpose view as far as possible.[2] If it falls short, in some appropriate respect, we might be able to construct a persuasive argument for a qualitative discontinuity in man's psychological evolution. In

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

our present ignorance of cognitive mechanisms, however, we cannot make such an argument convincing to anyone who is not already inclined to believe it. All we can do is to note questions that a general-purpose theory must answer.

For example, how is our unusual faculty for language related to our unusual intelligence? A general-purpose theorist might be tempted to argue that human language is a consequence of human intelligence, but that approach is too simple. Not all members of our intelligent species are intelligent, yet the rudiments of language are discernible even in unfortunates barely able to perform their vital functions. You do not have to be smart in order to talk; you have to be human. If human language and human intelligence are to be related in some biologically valid theory, language must depend on the kind, not the amount, of human intelligence.[3] But what do we know about different kinds of intelligence?

Intelligence is said to be the ability to profit from experience. It must be possible to collect information about the environment, which psychologists call perception; it must be possible to retain that information, which psychologists call memory; it must be possible to prefer some state of affairs to others, which psychologists call motivation; it must be possible to modify behavior on the basis of remembered information in order to increase the probability of achieving preferred states of affairs, which psychologists call learning. Perception, memory, motivation, learning are innate general-purpose cognitive mechanisms that have been studied in many species other than man. To which of these would a general-purpose theorist attribute the peculiar human faculty for language?

Whichever cognitive process one chooses, the question of general purpose versus special purpose persists. Take learning, for example. Many animals can learn, yet they do not learn to speak a human language. Is man simply able to learn more than any other animal, or has he some special linguistic mode of learning? Is there anything intrinsically different about, say, learning to find food and learning to speak English? Or can the same general principles of learning explain both? This argument can be repeated for any of the cognitive mechanisms one cares to explore.

I have said enough to indicate the background against which I view the two significant conceptual advances that I want to discuss. Let me turn now to the advances themselves: transformational generative grammar and the simulation of cognitive processes by computers. Transformational generative grammar is a theory of language that can be attributed to a single originator, Noam Chomsky.[4] No one individual played a similarly decisive role in the use of computers to simulate intelligent human behavior, although A. M. Turing probably deserves credit for setting the problem.[5] Both advances appeared in the 1950s, both were elaborated and extended in the 1960s, and both have affected psychological theories of language in the 1970s. Both emerged from conceptual advances in logic and metamathematics, and both are far too technical for any adequate explication in a few pages. I will say only enough about each to support what I see as their effects on psychological theory in general, and on psychological theories of language in particular.

First, the grammar. Although most of the time of any serious grammarian is spent in the detailed study of some particular language, the conceptual

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

advance of interest to me concerns the general theory of language. I shall, therefore, concentrate on formal and abstract aspects of Chomsky's thinking, aspects that are basic to, but usually implicit in, the actual work of constructing transformational generative grammars for English, French, Russian, or other specific languages.

Suppose that we are given a finite vocabulary of symbols and asked to consider the infinite set of different strings that is generated by arranging those symbols in all possible sequences of finite length. Any device that selects some subset of those strings is a grammar, and the strings it selects is a language. For example, if the vocabulary is all the words of English (including all inflectional, derivational, and compound forms), then a grammar of English is a selective device—a device that selects, from all the possible sequences of words, all and only those strings that are grammatical sentences. The English language is the set of strings so selected.

The substantive question, of course, is how to characterize the grammar—the device that selects all and only the grammatical strings. But note first the rigorous structuralism of this definition of a language. It tells us what English is, but provides no hint of its function, of what it might be used for.

Subsets of the set of all possible strings can be selected in an uncountably infinite variety of ways. For example, the possible strings might be ordered and a coin flipped successively for each string: heads would make it grammatical, tails ungrammatical. Such a random selective device would generate a set of grammatical strings that would be a language, in this formal sense, but it would be nothing like a natural language. We know it would be unnatural because any person who wanted to learn it would have to memorize each grammatical string separately; no rules or generalizations would help him learn this random subset. Because there would be an infinite number of grammatical strings to memorize, the task could never be mastered completely. And because children do normally master a natural language in a finite childhood, natural languages must follow intelligible rules, not random choices. We can, therefore, limit our consideration of selective devices for natural languages to those that can be formulated in terms of a finite (and probably relatively small) number of rules.

This approach helps us to understand why grammarians have always talked about rules, but the kind of rules they have traditionally talked about are ill suited to the design of a selective device. Prescriptive rules (e.g., do not split infinitives) define a style, not a language, and so must be ignored, but even the valid descriptive rules of traditional grammar (e.g., subject and verb must agree in number) have to be reformulated to make their presuppositions explicit. To say, for example, that every sentence must have a subject phrase and a predicate phrase is true enough, but it is too abstract. Before a selective device could use such a rule to recognize sentences, it would have to have rules to recognize phrases, and to tell which were subject phrases and which predicate phrases. Obviously, the new rules must be much more detailed and explicit than were those required for the purposes of traditional grammar, because they must be usable by an abstract selective device totally lacking any human faculty or intuition for language.

Perhaps the first kind of selective device one thinks of would begin at the left end of a string (for English) and examine each symbol successively. After all,

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

that is how sentences are spoken and how the words present themselves to a listener. Such a device would have, for every admissible beginning, rules determining admissible continuations. If the device reached the end of a string without any of those rules being violated, it would accept the string as a grammatical sentence; otherwise, the string would be rejected.

We should pause to note that a selective device (grammar) consisting of rules for proceeding from left to right (in English) through a string of words would also be compatible with the kind of theories that dominated psychological thinking about learning in the 1950s. Those theories were constructed to describe the formation of bonds between stimuli (S) and responses (R), and long strings of responses were thought of as chains of such S–R bonds. For example, some environmental stimulus might lead to the response of uttering a word; that word, once uttered, would provide a stimulus for uttering the next word-response; the next word would then be a stimulus for another word-response; and so on.[6] If one could state the rules of grammar in terms of the possible continuations of a string of word-responses, those rules could be interpreted directly as the product of S–R learning. Psychologists interested in reducing language learning to general processes involved in all learning have invested considerable ingenuity in elaborating such ideas.

In spite of their plausibility, however, Chomsky showed that such devices are inadequate as grammars of natural languages.[7] In their simplest form they do not account for the phrase structure of sentences, and therefore they cannot account for sentences in which a phrase of indefinite length is inserted into another phrase: "The men who . . . are here" must preserve number agreement between *men* and *are* over indefinitely long sequences of intervening words that can be inserted after *who*. A selective device that cannot recognize phrase structure must have separate rules for each intervening grammatical substring and, because the intervening substring can be indefinitely long, that means it must have indefinitely many rules. But requiring a learner to memorize indefinitely many rules for admissible continuations is no improvement over memorizing indefinitely many sentences.

At the very least, therefore, the selective device must be able to recognize constituent phrases in more complex grammatical phrases. Chomsky considered next the proposal that a selective device for natural languages might be constructed in terms of "rewriting rules" that can characterize phrase structure. For example, it might have a rule that rewrites Sentence as Noun Phrase followed by Verb Phrase. Then another rule would rewrite Noun Phrase as, say, Article followed by Noun, and another would rewrite Verb Phrase as Verb followed by Noun Phrase. And there might be terminal rules that rewrite Article as *the* or *a*, Noun as *boy*, *doll*, *dog*, or whatever. Such devices can generate grammatical strings of words and assign to each string its appropriate phrase structure.

As just described, this generative device sounds like a theory of how people produce the sentences they utter, an interpretation that has caused considerable confusion. But Chomsky was still thinking of it as a selective device. Given any arbitrary string, the device would accept it as grammatical if it could generate that string by following its rewriting rules. *How* the device was to determine whether or not it could match the input string was not explained. A great

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

variety of ways to do it could be imagined; which particular way people do it was left open as an empirical question. Once again, note the structural emphasis—this time on the characterization of sentence structure in the absence of any explanation of how such structures are produced by speakers or interpreted by listeners.

Although systems of rewriting rules are adequate to characterize the phrase structures of simple sentences, Chomsky concluded that they are not adequate for complex sentences. Or, if they are adequate in some superficial sense, they are not adequate to capture many valid linguistic insights into relations among sentences. For example, "The men are here" is intuitively related to "Are the men here?" yet a grammar having only rewriting rules would generate them differently (generate in Chomsky's sense) and would miss the generalization. He therefore equipped his grammars with another kind of rule, the grammatical transformation, that could take the phrase structure underlying "The men are here" and transform it into the phrase structure underlying "Are the men here?" A transformational grammar not only stated the relation between the sentences explicitly—it also permitted a great economy in the statement of the rewriting rules. The proper division of theoretical labor between rewriting and transformational rules (in particular, how to limit the power of the transformational rules in intuitively reasonable ways) has been a central concern of those who have worked with this conception of grammar.

Once again, however, confusion resulted when psychologists assumed that Chomsky was making claims about how language is used—about the cognitive processes involved in producing or understanding sentences. If the interrogative sentence "Are the men here?" is derived transformationally from the phrase structure underlying the declarative sentence "The men are here," it would seem to imply that a speaker first thinks of the declarative sentence and then transforms it into the interrogative form, and that the listener has to reverse this transformation in order to recover the declarative form before he can understand the question. Because I took this line myself in the early 1960s, I am responsible for more than my share of this confusion. In self-defense, I believe that my explorations forced Chomsky to clarify his claims, or at least state them more explicitly.[8] It is now clearly on the record that he intended to describe the structure of language, not its manner of production or interpretation in actual use.

I believe that it is instructive to consider *why* psychologists have so persistently misunderstood Chomsky's structuralism. The confusion was most prevalent among American psychologists who, since the early days of this century, have taken for granted a functional approach to the mind. The roots of American functional psychology were in the pragmatic philosophy of Pierce, James, Dewey, and Mead, and it fed on the evolutionary theories of Spencer, Darwin, and Galton.[9] It suited the American temperament to be more interested in what minds are good for than in how they are constructed. What they are good for, of course, is to guide behavior, and this line of thought, when carried to its ultimate conclusion, led to the American school of psychology known as behaviorism. But even those psychologists who did not reject the mind in favor of purely behavioral analysis still agreed that an organism's adjustment to its environment was the proper subject matter for their science.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

American functional psychology dealt with mind in use, and its goal was to describe the processes whereby an organism came to terms with its world.

Functionalism is so deeply ingrained in American psychological thought that many of my psychological colleagues still believe that either Chomsky is making claims about cognitive processes or else he is saying nothing of psychological significance. When Chomsky demonstrated that the simple application of S–R learning theory to chained responses is inadequate in principle as an account of grammar, he could not be ignored. But confusion resulted when psychologists assumed that Chomsky was describing another kind of cognitive process that could be substituted for S–R chaining in a functional account. In fact, however, Chomsky's thought did not follow the traditional lines of functionalism. Chomsky did not deal with language in use, but instead he revived a kind of structural theory foreign to the prevailing belief that all valid psychological explanation must be stated in terms of processes adequate to account for various functions performed by organisms in the course of adapting to their environments.

Of course, if Chomsky's ideas had been either wrong or without psychological implications, psychologists would not have been forced to cope with them; his structuralism might have remained a local aberration in linguistics. But those of us who embraced them found our thinking about language, and about all the mental processes that language supports, radically changed. Instead of asking how S–R chains are learned, we began to ask how rules are learned, and how they support generalization from the finite set of sentences we have actually heard to the infinite set of sentences we are able to understand on first hearing. It changed completely our ideas of what a child is learning when he masters his native tongue, and how that learning process should be studied. It changed our ideas of how the brain might work, and of the evolutionary preparation for this unique mode of communication. Most of us held fast to our functional predilections, and so were embarrassed by the lack of any persuasive account of the cognitive processes involved, but the kinds of processes we were looking for in the light of Chomsky's work were entirely different from those we had expected to find before.

Let me turn now to the second conceptual advance, the simulation of cognitive processes. Here we must begin with the assumption that any function that can be computed can be computed by a general-purpose computer. If, for example, some set of rules enables a person to compute his moves in a chess game, we need merely specify those rules precisely in order to program a computer to make the same computations, that is to say, in order to program a computer to simulate the person's chess-playing behavior. It is generally assumed that if a computer's behavior in all chess situations is indistinguishable from a person's, then whatever we mean when we say the person was thinking about those situations is the same thing we mean when we say the computer was thinking about them.

The catch is that we have to specify precisely the rules that the person was following. It is easy enough to incorporate the rules that define the legal moves of chess, because those have been stated precisely. To simulate a chess player, however, we must also incorporate the rules that he uses to select among the legal possibilities, and those rules have not been stated precisely. Lacking any

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

way to determine a player's rules of strategy from examination of his brain tissues, we can do little more than ask him to think aloud as he plays. Then we can try to devise rules that will, when converted into computer programs, cause a computer to consider the same alternatives in the same order and reach the same decisions. This approach is most successful with players of moderate skill. A novice will not yet have developed consistent strategies of any interest; grandmasters have internalized their strategies so thoroughly that they are no longer accessible to introspective report.

This line of research was pursued most vigorously by Allen Newell, Herbert Simon, and their students, who, after considering a wide variety of different reasoning tasks, evolved a general theory of human problem-solving expressed in the form of computer programs.[10] The theory includes the representation of the task environment, the definition of goals and subgoals, the evaluation of alternative states of affairs in terms of those goals, and the order in which alternative courses of action will be explored. The plausibility of the theory derives not only from the accuracy of the simulations it supports, but also from its similarity to the various accounts of problem-solving strategies developed by psychologists, mathematicians, and other scientists on quite different grounds. The advantage of the theory is its explicitness, because nothing can be left undefined in stating the processes to be performed by the computer. Never before were psychological theories of reasoning stated in such concrete and explicit detail.

One presumes that the information-processing performed by the computer parallels, at least at a gross level, the cognitive processes of a human problem solver. At a fine level, of course, the parallel fails; whatever else a brain may be, it is certainly not a high-speed digital computer. Precisely how far down into the details of the computer program the parallel can be maintained is an open question on which opinions differ depending on one's enthusiasm for this kind of theory construction. The theory is, of course, severely cognitive; psychologists impressed by the emotional component in human thinking are inclined to be critical, but the force of this criticism remains to be assessed.

The availability of a computer that can think like a person is suggestive of many practical applications, not the least of which is that persons can be replaced by such computers in many jobs that required problem-solving hitherto beyond a computer's grasp. Leaving aside the troublesome question of whether this development would increase unemployment or free people for more creative work, we can still ask whether such applications would be maximally efficient. It has generally turned out that when some routine information-processing task performed by human beings has been sufficiently well analyzed to permit simulation by computers, there is usually a more efficient way for computers to do it. To adopt an often used example, it would be possible to simulate human walking by machine, but a walking machine would not be the most efficient mode of transportation. Because efficiency receives higher priority in practical applications than does similitude, most of the seminal work in this field now goes under the name of artificial intelligence, rather than the simulation of cognitive processes.[11] For workers in artificial intelligence, no special status attaches to the particular way people would perform a function; the human way is but one among many, and usually not the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

best for particular applications of limited generality. Obviously, the fact that human beings can perform a particular function at all is an existence proof that there is some way to do it, but whenever there is one way, there are usually indefinitely many other ways as well.

I believe that the major import of this work for psychology has not been the theory of problem-solving developed by Newell and Simon, valuable though it is, but rather a general freeing of the scientific psychologist's imagination. Prior to the computer age, psychologists who wanted to formulate their theories in terms of plausible mechanisms that did not conceal more mentalistic assumptions than they explained were limited to such physiological mechanisms as had been revealed by studies of nerve conduction, simple reflexes, or conditioning. In those days, for example, to say that a person's behavior was the outcome of his evaluation of several imagined alternatives for achieving a particular goal would have been considered the kind of mentalistic description that a good scientist was expected to reduce to S–R connections. Today a psychologist can, with no loss of rigor or plausibility, sketch such theories in terms of flowcharts for the information-processing involved, where every box in the chart represents a process that can be performed mechanically. Those of us who have lived through the transition recognize that scientific theories in psychology are now much more interesting—and probably much closer to what psychology ought to be. At least the new theories do not have to reduce everything to conditioned reflexes—which, even if it were possible, would be about as informative as saying that a cathedral is constructed of stones. The subtle details of information-processing offer a richer account, and the complexity of psychological functions that can be considered is much greater.

Unlike workers in artificial intelligence, psychologists inspired by the computer metaphor do have a special interest in how people perform a function. The spirit of simulation can be seen in hundreds of published accounts phrased in terms of computational algorithms or heuristics, and more articles on the subject appear in every issue of the journals of experimental psychology. There is no conflict here with artificial intelligence—the first task is to conceive *any* way that a function might be performed, and on that score collaboration can be complete. Given one solution, however, alternatives can then be developed to meet the different criteria: similitude or efficiency.

A psychological model of a language user constructed along these lines must cope simultaneously with a great variety of processes that a linguist would distinguish and analyze separately. The model must recognize words, which means that some decision must be made about how lexical information is stored in memory (which generally holds implications for how new lexical information is added to memory); it must be able to parse sentences—simple sentences, anyhow—into their phrasal constituents, which requires a translation of grammar into mechanical processes; it must be able to compute implications of the statements that it analyzes; it must make the statement and its implications available to the executive program that governs the model's strategy in achieving its goals; it must formulate appropriate sentences of its own to communicate or request information. None of these functions is trivial; the attempt to build a language user can be seen as setting a whole range of specific psychological processes in a general context where their separate responsibilities and mutual dependencies can be seen in a new light. The result has been to lift psychologi-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

cal theories of cognitive functioning to significantly higher levels of integration and sophistication.

The computer metaphor is not without its critics, however. Those psychologists who view consciousness as the constitutive problem of psychology—and they would include all those who try to analyze unconscious processes, for there is no unconsciousness without consciousness—are particularly likely to see the computer, for which we have no analogue of the conscious/unconscious distinction, as just another diversion from the real issue, not so much wrong as beside the point. They feel that it makes a difference whether a rule is known consciously or unconsciously, explicitly or implicitly, but how that difference should be represented in a computer model is discouragingly unclear. However, those psychologists who appreciate the deeper understanding that computer models have made possible have not been deflected by such doubts. An answer to one question is not wrong simply because it does not answer all questions.

One reason so many serious psychologists have accepted the computer simulation as their paradigm for scientific explanation is that it agrees so well with their functional predilections. Programs defining a sequence of information processing must have a structure, of course; their degrees of freedom in combining processes are strictly limited. But they are characteristically goal-directed, with success defined in terms of the ability to perform specific operations required to achieve the goal. Thus, they stand in sharp contrast to the kind of structural theories put forward by Chomsky and other transformational generative grammarians.

This contrast can blossom into conflict when information-processing techniques are applied to spoken or written language. This is not the place and I am not the person to review all the various attempts to process natural language by machine—analysis and synthesis of speech, automatic translation, question-answering, computer-assisted instruction, content analysis, and the like. I can illustrate what happens in very general terms, however, by considering applications that involve a first stage that takes information coded in the sentences of a natural language and recodes it into routines that will control the operation of a computer. In computer terminology, such a device is called a compiler: the grammatical sentence is treated as if it were a program written in some high-level programming language that must be translated into the operational code of the computer.[12] If, for example, the input sentence is "Are the men here?" the compiler is expected to translate it into a routine that will search the information available to the computer and then formulate an answer in another sentence that can be given in response: "Yes," "No," "What men do you mean?" or whatever. One step in the compilation is to parse the sentence—to determine its phrase structure—but how this function is performed may bear only a superficial resemblance to the structural accounts formulated by transformational generative grammarians. For the computer application, the important thing is to be able to provide an answer. If answering does not require a full transformational analysis, then so much the better.

All this is very well for questions, requests, or commands, but what would a computer do with "The men are here"? In a grammatical analysis, this sentence would have one relatively simple description. In a computer program, however, it might be treated in several different ways, depending on the goals that the computer was programmed to achieve. It might, for example, be compiled as a

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

routine to add this information to its data base. Or it might be compiled as a routine to test its truth in the light of other information available to the computer. Or it might be compiled as a routine to initiate some preset pattern of action, or to ask the men's names, or even to ignore it as irrelevant. What the computer will do will depend on far more than the grammatical structure of the sentence.

To a functionally oriented psychologist, the computer system seems a much more reasonable model for psychological explanations than does the grammatical analysis. On the other hand, years of struggling with grammatical subtleties make it a simple exercise for a grammarian to find constructions that the computer will analyze incorrectly, regardless of what goals it was intended to achieve subsequently. To a structurally oriented linguist, therefore, the computer system seems at least inadequate, if not downright wrong. The question of whether there is something wrong with the computer system in principle, or whether its deficiencies could be corrected by better programming, has generated more argument than answers. Meanwhile, those of us who would like to bring the structural and the processing accounts together live in constant disagreement with colleagues who see them as distinct as are, say, the economics and the physiology of dairy cows.

Chomsky has proposed one way to live peacefully with both approaches without rejecting either.[13] Transformational generative grammar, he says, is a theory of the ideal language user's linguistic competence, of what he knows about his language. The psychologist is interested in performance, in how that knowledge is put to use. A psychological theory of performance must incorporate a linguistic theory of competence, but exactly how is left as an exercise for the psychologist. This distinction between competence and performance was snapped up gratefully by perplexed psychologists, but unfortunately it has been interpreted in a dozen different ways, with little consequent reduction in perplexity.[14]

The larger philosophical issue raised by this distinction, of course, concerns the place of knowledge in a theory of action. Because there is no hope that so complex a matter could be settled from the armchair, little is gained by identifying it in these terms. Moreover, in the linguistic version of this problem, the implicit character of most grammatical knowledge is enormously complicating. What does it mean to say, for example, that a person *knows* some complex transformation for deriving the structure of sentences containing restrictive relative clauses? It does not mean that he can state the rule, for only trained grammarians can do that, and a foreigner who had memorized the rule would not necessarily be able to follow it in his own speech. It does mean that, under ideal conditions, his own sentences will conform to the rule, and that he can identify deviations as mistakes. But it must mean something more than that. A clever dog that can catch a ball in its mouth would not be thought to *know*, even implicitly, the differential equations of motion, even though his behavior in extrapolating the path of the ball conforms to those equations, and a moving object that deviated from them would be treated differently. How are we to characterize knowledge that is somehow intermediate between conscious expressibility and mere conformity? A psychologist can only hope that his own work need not await the successful resolution of such enigmas.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Personally, I feel that much of the tension and perplexity arise not from different interests in knowledge or action so much as from differences between structural and functional theories. As matters stand presently, psychologists are trying to combine a structural account of competence with a functional account of performance. The whole problem might look very different if, say, we considered a functional theory of grammar or a structural theory of mind. No principled argument ensures that such theories would be absurd, and plausible instances are not difficult to find. In linguistics M. A. K. Halliday,[15] for example, has proposed a functional grammar that has been incorporated in at least one computer system for processing language. And in psychology Jean Piaget[16] has proposed structural descriptions of successive stages in cognitive development. These theories have enjoyed little popularity in America; apparently it is as difficult to integrate function into structural theories of linguistics, or structure into functional theories of psychology, as it is to integrate a structural theory of linguistics with a functional theory of psychology. As long as this difference remains unresolved, the current state of tension will persist, for neither school of thought is likely to abandon an approach that has contributed so much and holds such promise for the future.

What, then, can the facts of linguistic communication teach us about the nature of the human mind? Not nearly all we would like to know. Structural theorists of grammar find uniquely complex knowledge structures; functional theorists of language in use fail to find uniquely human cognitive processes. There the matter rests. Before psychologists can exploit fully the implications of both of the conceptual advances I have tried to characterize, they will have to come to grips with the fundamental theoretical complementarity between them.

References

[1]S. Hook (ed.), *Language and Philosophy* (New York: New York University Press, 1969).

[2]G. A. Miller, "Four Philosophical Problems of Psycholinguistics." *Philosophy of Science* 37 (1970): 183–199.

[3]E. H. Lenneberg, *Biological Foundations of Language* (New York: Wiley, 1967).

[4]N. Chomsky, *Syntactic Structures* (The Hague: Mouton, 1957).

[5]A. M. Turing, "Computing Machinery and Intelligence." *Mind* 59 (1950): 433–460.

[6]G. A. Miller and J. A. Selfridge. "Verbal Context and the Recall of Meaningful Material." *American Journal of Psychology* 63 (1950): 176–185.

[7]N. Chomsky, "On Certain Formal Properties of Grammars." *Information and Control* 2 (1959): 137–167.

[8]N. Chomsky, *Aspects of the Theory of Syntax* (Cambridge, Mass.: The M.I.T. Press, 1965).

[9]E. G. Boring, *A History of Experimental Psychology* (2nd ed., New York: Appleton-Century-Crofts, 1950).

[10]A. Newell and H. A. Simon, *Human Problem Solving* (Engewood Cliffs, N.J.: Prentice-Hall, 1972).

[11]M. Minsky, "Steps Toward Artificial Intelligence," in E. A. Feigenbaum and J. Feldman (eds.), *Computers and Thought* (New York: McGraw-Hill, 1963).

[12]D. J. M. Davies and S. D. Isard, "Utterances as Programs," in D. Michie (ed.), *Machine Intelligence 7*. Edinburgh: Edinburgh University Press, 1972.

[13]Chomsky, *Aspects of the Theory of Syntax*.

[14]G. A. Miller, "Some Comments on Competence and Performance," in D. Aaronson and R. W. Rieber (eds.), *Developmental Psycholinguistics and Communication Disorders* (New York: New York Academy of Sciences, 1975).

[15]M. A. K. Halliday, *Explorations in the Functions of Language* (London: Edward Arnold, 1973).

[16]J. Piaget, *Structuralism* (New York: Basic Books, 1970). Translated by Chaninah Maschler (original French edition 1968).

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

EDWARD O. WILSON

Biology and the Social Sciences

The opposite of a correct statement is a false statement. But the opposite of a profound truth may well be another profound truth.—Niels Bohr

FOR EVERY DISCIPLINE in its early stages of development there exists an antidiscipline. For many-body physics, particle physics; for chemistry, many-body physics; for molecular biology, chemistry; for cellular biology, molecular biology; and so forth. With the word *antidiscipline* I wish to emphasize the special adversary relation that exists initially between the studies of adjacent levels of organization. This relationship is also creative, and with the passage of a great deal of time it becomes fully complementary.

In this article I will argue that biology has now moved close enough to the social sciences to become their antidiscipline. Hitherto biology has affected the social sciences largely through technological manifestations such as the benefits of medicine, the mixed blessings of genetics, and the specter of population growth. Although of great practical importance, these matters are wholly trivial with reference to the conceptual foundation of the social sciences. The conventional academic treatments of "social biology" and "social issues of biology" present some formidable intellectual challenges, but they are not concerned with the core of social theory. Many scholars judge this core to be the deep structure of human nature, an essentially biological phenomenon that is also the primary focus of the humanities.

If it is true that biology is the antidiscipline of the social sciences, the past failure of the social sciences to develop a common body of theory is understandable. The reason is that the relevant branches of biology—neurobiology and sociobiology—are only now becoming mature enough to attain a juncture with the social sciences. Although it would be premature to say that biology can revolutionize the social sciences, at least to the extent that chemistry has revolutionized biology, it would be equally premature to say that it cannot.

This proposition is well worth examining. To that end I will first describe precedents from the history of biology, then turn to some developments within the individual social sciences that suggest a growing susceptibility to biological explanation. Finally, acknowledging my obligation as a student of the antidiscipline, I will attempt to define the obviously strong limitations of biological reductionism.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Disciplines and Antidisciplines

In general, the practitioners of a given discipline in its early, natural-history phase are concerned with the discovery and classification of phenomena. They stress novelty and particularity. In terms of the classic thematic dualities of science,[1] their explanations are characteristically holistic, emphasizing pattern and form over units of construction. In the early phase, specialists are also likely to be dualistic in philosophy, questioning whether their phenomena are directly subject to the laws of the remainder of science.[2] And in later stages, having been converted, they are still more concerned with what Victor Weisskopf has called extensive as opposed to intensive research, the use of existing theory to explain the widest possible range of phenomena as opposed to the search for fundamental laws.[3]

Members of the antidiscipline are likely to be monistic with reference to the discipline and dualistic with reference to their own subject. Having chosen as their primary subject the units of the lower of the paired levels of organization, they believe that the next discipline above can be reformulated by their laws. Their interest is relatively narrow, abstract, and exploitative, lacking the totemic attachment to phenomenology displayed by the most devoted students of the discipline above. Thus P. A. M. Dirac, speaking of the theory of the hydrogen atom, could say that its consequences would unfold as mere chemistry, whereas the biochemist Franz Hofmeister responded to the recent great advances in cell structure by recommending (in 1901) that biologists pay more attention to enzymes.

It is easy to see why each discipline is also an antidiscipline. A tense creative interplay is inevitable because the devotees of adjacent levels of organization are committed to different methodologies when they focus on the upper level. By today's standards a broad scholar can be defined as one who is a student of three subjects: his discipline, the lower antidiscipline, and the subject to which his specialty stands as antidiscipline. A well-rounded cellular neurophysiologist, for example, is deeply involved in the microstructure and behavior of single cells, but he also understands the molecular basis of electrical and chemical transmission, and he hopes to explain enough of neuron systems to help account for the more elementary patterns of animal behavior.[4]

Cell Biology and Biochemistry

In the late 1800s cell biology and biochemistry grew at an accelerating pace. Their relationship during this period was very complicated, but it can be broadly characterized as fitting the schema just described. The cytologists were excited by the mounting evidences of an intricate cell architecture. They had also deduced the mysterious choreography of the chromosomes during cell division, setting the stage for the emergence of modern genetics and experimental developmental biology. Many biochemists, on the other hand, remained skeptical of the idea that so much structure exists. They emphasized the possibility of artifact production by the chemical reactions used in cytological preparations and stood apart from the debate then raging over whether protoplasm is homogeneous, reticular, granular, or foamlike. Their interest lay in the more "funda-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

mental" issues of the chemical nature of protoplasm, especially the newly formulated enzyme theory of life.

In general, biochemists judged the cytologists too ignorant of chemistry to grasp the basic processes, whereas the cytologists considered the methods of the chemists inadequate to characterize the structures that diagnose the living cell. The renewal of Mendelian genetics and subsequent progress in chromosome mapping did little at first to effect a synthesis. Biochemists, seeing no immediate way to encompass classical genetics, by and large ignored it.

Both sides were essentially correct. Biochemistry has now explained so much of the cellular machinery by its own terms as to justify its most extravagant early claims. But in achieving this feat (mostly during the past thirty years) it was partially transformed into the new discipline of molecular biology—biochemistry that entails particular spatial arrangements and movements of large molecules. Cytology forced the development of a special kind of chemistry and the employment of a wide array of powerful new methods, including electrophoresis, chromatography, density-gradient centrifugation, and X-ray crystallography. At the same time cytology metamorphosed into modern cellular biology. Aided by electron microscopy, it converged in language and outlook toward molecular biology. Finally, classical genetics, by switching from *Drosophila* to the ultrafast-breeding and far more simply constructed bacteria and viruses, has incorporated much of biochemistry to become molecular genetics.

Progress over a large part of biology was fueled by competition among the various attitudes and themata derived from biology and chemistry—the discipline and its antidiscipline. Joseph Fruton, a biochemist who has paid close attention to this Hegelian interplay, has suggested that inevitably "such competition is attended by tensions among the participants. I venture to suggest that this competition and these tensions are the principal source of the vitality of biochemistry and are likely to lead to unexpected and exciting novelties in the future, as they have in the past."[5]

Ecology and Population Biology

Modern ecology has had a troubled history. As recently as ten years ago it was painfully unfashionable in many American universities. An anecdote will illustrate how serious the situation had become. The Department of Biology at Harvard University was at that time increasingly dominated by molecular and cellular biology, and there appeared to be little chance of adding new faculty members specializing on populations or ecosystems. One afternoon I proposed departmental membership for a distinguished ecologist who had been appointed by another school within the university. One of the molecular geneticists, a new Nobel laureate, said, "Are they out of their minds?" When I asked what he meant, he responded that anyone who hired an ecologist must be out of his mind. After the meeting one of my senior colleagues in evolutionary biology suggested that I not refer explicitly to ecology in the future, because it had become a "dirty word." The ecologist was invited much later, after the molecular and cellular biologists had formed independent administrative units.

Ecology is almost embarrassingly fashionable now, and I should add at once that the several ecologists subsequently added to the Department of Biology are

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

on very cordial terms with all their colleagues. But the lesson learned at Harvard and at other universities with similar experiences was, I believe, that few scientists are willing to share resources with those whose research is more than one level of organization removed. Populations and communities, the central concern of ecology, are separated from molecules and cells by an entire level, that of the organism. In an environment ruled by competitive research, profit can be extracted only from the discipline and antidiscipline.

The problem with ecology, however, was more than lack of sympathy from molecular and cellular biologists. The difficulty came ultimately from ecology's focus on the highest level of biological organization, the community, and the weakness of its connections to fundamental population biology. Without quite realizing it, ecology was an orphan discipline. By the 1950s theoretical population genetics, one of the logical antidisciplines, had grown into a technically formidable, even arcane subject. Its models were derived from the distinctive chromosomal mechanics of Mendelian heredity, and it dealt almost exclusively with the interplay of the Darwinian operators—selection, mutation pressure, gene flow, genetic drift, and meiotic drive. Ecology was reduced to the single parameter of the selection coefficient. By a remarkable piece of bad luck, experiments and field studies were concentrated on *Drosophila*. Fruit flies are superb insects for rapid genetic and demographic analysis in the laboratory, but they are singularly hard to find and to study in their natural habitats. Much of what is known of their ecology was deduced from their appearance at food traps. Partly as a result of this historical accident, population genetics remained apart from ecology.[6]

Similarly, with the single exception of pure demography, the fundamental theory of population ecology had not advanced much beyond the principles advanced by Alfred Lotka, Vito Volterra, and R. A. Fisher in the 1920s and early 1930s. Ecology, consisting largely of analysis at the level of entire communities, developed as an ever more elaborate descriptive science. In some of its branches, such as biogeography and phytosociology, the systems of classification and quantitative description reached phantasmagoric extremes.

The time was obviously ripe for a new effort that would bring population genetics and population ecology together and cast them in their proper role as antidisciplines to community ecology. In the late 1950s and early 1960s a small group of younger population biologists (I confess to being a member) set about self-consciously to make such an effort. Several meetings were held to discuss the matter and to some extent to divide the labor. The undoubted leader was the late Robert H. MacArthur, whose work on complex community phenomena was exceptional in its originality and discrimination of important problems.[7] MacArthur's seminal contribution was his 1957 analysis of the population-abundance frequency curve of a community of bird species.[8] Descriptive ecologists had accumulated a large amount of information on this subject but had done little more than fit empirical curves to it. MacArthur used the data to test a series of competing hypotheses based on models of various forms of interaction among species populations, a process which up to that time had been very difficult to study directly in nature. Although his method has had only limited success when applied to other biological communities, it validated postulational-deductive theory and demonstrated that leaps of the imagination can lay open the most complex ecological processes and give them new meaning.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

The response of other ecologists to such renewed model building at the population level was predictably mixed. Many of those devoted to painstaking descriptive work refused to believe that general laws could be so cheaply bought. The patterns they discerned seemed too elaborate, the variables too numerous, and the mark of history too deep and idiosyncratic to make general models anything more than a clever illusion. A community of organisms is a tangled bank, to use Darwin's famous phrase in the closing paragraph of the *Origin of Species*, or a uniquely woven tapestry, in the metaphor of one of the recent critics. Prominent ecological schools still exist, comprised of ecosystems analysts, theoretical population ecologists, physiological ecologists, and so forth. It is cytology versus biochemistry all over again. But the history of this subject, if I have interpreted it correctly, is now entering the middle phase of the classic ontogeny. Broad areas of agreement are already apparent, and model building at the population level has become routine.[9]

The Social Sciences and Biology

Let us now return to the original proposition that biology is the antidiscipline of the social sciences. This assessment is not congenial to the prevailing view in the social sciences and humanities that human social life is the nearly exclusive product of cultural determinism, constrained only by the most elementary and unstructured biological drives. There is a strong tendency to think of our own species as entirely plastic and hence all but equipotent in the design of its social institutions. However, this conception will not stand close scrutiny. A comparison of the literally tens of thousands of other highly social species on Earth, from colonial coelenterates through the social insects to the most social of the birds and mammals, reveals that the sum of all the varieties of human social behavior occupies only a small envelope in the space of realized social arrangements.[10]

Anthropocentrism is a disabling vice of the intellect. I am reminded of the clever way Robert Nozick deflated our sense of superiority to other animal species in order to make his principal argument for vegetarianism. If visitors from another planet happened to be far more intelligent and sensitive than ourselves, and applied our own criteria of relativity, they could proceed in good conscience to eat us.[11] By the same token, and to our considerable chagrin, scientists among them might find us uninteresting as a social species—just another cultural-linguistic variant on the basic mammalian theme—and instead turn to study the more theoretically challenging societies of ants and termites.

It is this quality of specificity and restriction that biologists have in mind when they speak of genetic determinism. In order to define a genetic trait precisely it is necessary to compare two or more states of the same character. To say that blue eyes are inherited, without further qualification, is not meaningful, because everyone knows that blue eyes are the product of the interaction of genes and the largely physiological environment that brought final coloration to the irides. But to say that the difference between blue and brown eyes is based partly or wholly upon differences in genes is a meaningful statement by virtue of being testable and translatable into the laws of genetics.

Human social behavior can be evaluated in the same way, first by comparison with the systems of other species and then, with far greater difficulty

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

and ambiguity, by studies of variation within the species. For example, certain general traits are shared with most other Old World primates, including size of intimate social groups on the order of 10–100; males larger than females, probably in relation to polygyny; a long period of socialization in the young, shifting in focus from the mother to age- and sex-peer groups; and social play strongly developed, with emphasis on role practice, mock aggression, and exploration. It is virtually inconceivable that primates, including human beings, could be socialized into the radically different repertories of insects, fish, birds, or antelopes; or that the reverse could be accomplished. Human beings, by conscious design, might well *imitate* such arrangements; but it would be a fiction played out on a stage, running counter to deep emotional responses and with no chance of persistence through as much as a single generation.

Homo sapiens is distinct from other primate species in ways that can be explained only as the result of a unique human genotype. Universal or near-universal traits include the facial expressions that denote basic emotions, and some other forms of paralinguistic communication; elaborate kinship rules that include incest avoidance; a semantic, symbolical language that develops in the young through a relatively strict timetable; close sexual and parent-offspring bonding; and others. Again, to socialize a human being out of such species-specific traits would be very difficult if not impossible, and almost certainly destructive to mental development. People might imitate the distinctive social arrangements of a white-handed gibbon or hamadryas baboon, but it seems extremely unlikely that human social systems could be stably reconstructed by such an effort.

It is significant that not only do human beings develop a species-characteristic set of social behaviors, but that these behaviors are generally mammalian, and most specifically Old-World primate in character. Furthermore, even the species-specific traits are logically derivable in some cases from the inferred ancestral modes still displayed by a few related species. For example, the facial expressions and some nonlinguistic vocalizations can be plausibly derived in phylogenetic reconstructions.[12] This is precisely the pattern to be expected if the human species was derived from Old-World primate ancestors (a fact) and still retains genetic constraints in the development of social behavior (a hypothesis).

An important quality of a genetic determinism is that it seldom entails the control of a single phenotype by a single gene. Polygenic inheritance is the rule, and the entity determined is not one trait but rather a range of possible phenotypes. For example, diabetes and schizophrenia possess moderate genetic components. The multiple genes underlying them produce a stronger tendency to develop the traits; they also prescribe the range of possible manifestations that are probable under specified environmental conditions. In a parallel way basic human social behaviors—including those as structured as male-group bonding, territoriality, and kinship rules—emerge as outermost phenotypes following behavioral development, the range and scope of which is constrained by the interaction of polygenes with the environment. With reference to this interaction, there is no reason to regard most forms of human social behavior as qualitatively different from physiological and nonsocial psychological traits.

Whatever the present social arrangements of our species, the biological foun-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

dation of human nature arose in populations that adapted to special environments. The prevailing hypothesis, which holds that the basic qualities were fashioned as an adaptation to a more predatory existence in open habitats, may or may not be correct in detail. The important point is that the emotional controls and the developmental pathways are considered to be structured in idiosyncratic ways that can be wholly understood only by retracing the ecological history of the species.[13] That such a relationship exists in other social species can be readily demonstrated.[14] This is why paleoanthropology, by reconstructing the Pleistocene African environment, has an important role to play in behavioral biology.

An unavoidable question is the extent to which social behavior varies genetically *within* the human species. This is a subject entirely removed from the distinctive properties of human behavior vis-à-vis that of other species. The evidence is strong that almost but probably not quite all differences among cultures are based on learning and socialization rather than on genes.[15] At the same time variation within populations is evidently great enough to create the potential for further human social evolution by population-wide genetic change. Studies comparing monozygotic twins with dizygotic same-sex twins suggests a genetic component in the variation of a large array of traits having an influence on the development of social behavior. These traits include verbal and number ability, word fluency, perceptual speed, memory, the timing of language acquisition, psychomotor skill, extroversion-introversion, homosexuality, and certain forms of neuroticism and psychosis.[16] Although not conclusive in themselves, such studies are strongly suggestive. Behavioral heritability is also enhanced by the undoubted existence of single, identifiable point mutations and chromosome aberrations, such as those causing the Lesch-Nyhan and Turner's syndromes, that alter various components of behavior differentially. Some geneticists have gone so far as to suggest that once the conditions make their appearance in spite of medical precautions, their study can permit the indirect genetic dissection of behavioral traits, in analogy to the technique used for nematodes and fruit flies.

My overall impression of the available information is that *Homo sapiens* is a typically animal species with reference to the quality and magnitude of the genetic diversity affecting behavior. If the comparison is correct, the psychic unity of humankind has been reduced in status from a dogma to a testable hypothesis. This is not an easy thing to say in the present political ambience of the United States, and it is regarded as punishable heresy in some sectors of the academic community. But the idea needs to be faced squarely if the social sciences are to be entirely honest. I cannot regard it as dangerous. Quite the contrary: the political consequences of its objective examination will be determined by our value system, not the reverse. It will be better for scientists to study the subject of genetic behavioral diversity than to maintain a conspiracy of silence out of good intentions and thereby default to ideologues.

Following this elementary excursion into genetic determinism, let us now consider how the various social sciences might be influenced by biological theory. In the brief sections to follow I have little more than scraps of information and impressions to offer. It is hoped that they will nevertheless suffice to provide a view of the reverse side of the social sciences as it has been glimpsed by a biologist approaching in that direction.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Anthropology

The central question of biological anthropology is the nature and strength of the coupling between cultural and biological evolution. Cultural evolution is Lamarckist and usually very fast; biological evolution is Darwinist and slower by at least an order of magnitude. Because the most rapid cultural changes track environmental fluctuations too brief in duration to influence directional genetic selection, cultural fitness can be expected to diverge frequently from genetic fitness.[17] But the divergence must be limited in degree and duration, because ultimately the newly created social environment will be tracked and the gap narrowed by natural selection. If this modification of basic sociobiological theory is correct, there should be two detectable consequences. First, it will be learning rules rather than specific cultural forms that are inherited.[18] And where the rules are most directly concerned with survival and reproduction, as in the case of sexual and parental bonding, incest avoidance, and xenophobia, they are likely to be the most rigid and structured. Marshall Sahlins has recently argued that the lack of clear correspondence between human kinship rules and the details of genetic kin selection theory disproves basic sociobiological theory as far as human beings are concerned.[19] But this goes much too far. Kinship rules are central to social organization in most societies, and in aggregate they appear to enhance inclusive genetic fitness. Richerson and Boyd[20] have suggested that four competing hypotheses can be posed for the explanation of kinship rules: detailed genetic control, rational strategizing, complete cultural determinism, and coupled cultural and genetic control. They conclude that the ethnographic facts are consistent only with the model of coupled cultural and genetic control.

The second expected consequence of coupling is that the genetic fitness conferred by particular cultural traits should be strongest in the oldest, culturally most stable societies. Thus it appears correct to focus attention on hunter-gatherer and persistent, preliterate herding and agricultural societies. This circumstance explains why anthropology has already become the social science closest to sociobiology. Explicit tests of sociobiological theory are being made in studies on polygyny, status, societal fissioning, territory, and warfare.[21]

It is probable that population biology will be simultaneously altered to accommodate the special problems of anthropology. In addition to the solution of the dual inheritance problem, there is a need for advances in the theory of group and kin selection to distinguish unilateral, "hard core" altruism from transactional, "soft-core" altruism. Also, the complexity of human population structures presents unique challenges to biology. Population boundaries are seldom sharp, often being confused by discordant linguistic, cultural, and historical-political patterns. Groups also shift rapidly in their loyalties, forming alliances in one year and dividing into quarreling factions the next. The present theory of population genetics and ecology is entirely inadequate to handle such complications.[22]

Psychology and Psychoanalytic Theory

Just as anthropology has been burdened in the past by the doctrine of complete cultural determinism, conventional psychology has been burdened by

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

general process learning theory. It was natural for Thorndike, Watson, and other pioneering psychologists to choose large animals, that is, birds and mammals, for their study objects, rather than insects and opisthobranch mollusks. Partly as a result of this choice, learning came to be accepted as the central process of behavior. Moreover, the learning process was viewed as being essentially equipotential: the same laws were theorized to apply to whatever learning process and organism are chosen. Thus Skinner could say in 1938, "The general topography of operant behavior is not important, because most if not all specific operants are conditioned. I suggest that the dynamic properties of operant behavior may be studied with a single reflex."[23] It was believed that by placing animals in simplified laboratory environments, where stimulation can be rigidly controlled, the most general laws would emerge.

This is a powerful idea, with seductive precedents in the physical sciences, and it has resulted in substantial advances in the study of animal and human behavior. Nevertheless, general process learning theory has started to crumble. In its place is appearing the description of a melange of specialized learning phenomena that conform to no general law except, perhaps, evolution by natural selection.[24] The full range of learning potential of each species appears to be separately programmed. According to the species, each animal is "prepared" to learn certain stimuli, counterprepared to learn others, and unprepared (neutral) with respect to still others. For example, adult herring gulls quickly learn to distinguish their newly hatched chicks but never their own eggs, which are nevertheless as visually diversified. Indigo buntings are prepared to learn the circumpolar constellations by which they orient their nocturnal migrations, but are counterprepared to learn other constellations. When chicks are shocked at their beaks while drinking water and simultaneously given a visual stimulus, they thereafter avoid the visual stimulus, but they do not learn to avoid an auditory stimulus presented the same way. The reverse is true when the shock is administered to the feet; that is, the chicks are prepared to learn sound but counterprepared to learn visual cues.[25] The timing of preparedness in the life cycle is also programmed and species-specific in ways that are readily interpreted as adaptations to the particular environments experienced by the species during their recent evolutionary pasts.

The hypothesis of learning rules as idiosyncratic evolutionary adaptations has seldom been examined with reference to the human species. It seems significant that phobias, which share some properties with imprinting in lower animals, are readily acquired against snakes, spiders, rats, and other potentially dangerous objects in mankind's ancient natural environment, but only rarely against such dangerous artifacts as knives and electrical outlets.[26] Language is acquired by small children through a progression of closely timed steps, involving distinctive vocalizations and phrase forms later replaced by adult language.[27] It is difficult to believe that the rules of this most human of all learning events have not been shaped by natural selection.

Psychoanalytic theory appears to be exceptionally compatible with sociobiological theory, a fact already appreciated by some of the psychoanalysts themselves.[28] If the essence of the Freudian revolution was that it gave structure to the unconscious, the logical role of sociobiology is to reconstruct the evolutionary history of that structure. When Freud speculated in *Totem and Taboo* on the primal father, primal horde, and the origins of the incest taboo, he created a

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

sociobiological hypothesis, but a poor one. The same is true of his insights into the conflict of self and society presented in *Civilization and Its Discontents*. Whether population biology and evolutionary theory can be used to restructure and objectively test some of psychoanalytic theory remains to be seen.[29]

Economics

Classic economic theory restricted itself to the goods and services that can be measured by money and market pricing. In recent years new, less easily quantified parameters have entered the equations, including time, human capital, and environmental quality.[30] A closer scrutiny is also being made of what Leibenstein has termed the X-efficiency factors, which include motivation, *esprit*, effort, persistence, and other psychological variables made mensurable.[31] In a word, microeconomics has begun to incorporate social psychology. It is now widely appreciated that human beings do not behave as rationalizing economic machines. Macroeconomic predictions of the future will almost certainly be based on the wiser perception of irrational elements in human nature.

To the extent that the new parameters of human irrationality are interpreted as an evolutionary product, the methods of economics will converge toward those of biology. Already, models in ecology and sociobiology have borrowed heavily from the graphical methods of economics. Optimization and decision theory are routinely used. The utility measure of biology is genetic fitness, and the enabling devices are anatomy, physiology, and behavior. I expect that once a method is developed for assessing the coupling of genetic and cultural evolution, the utility measures of economics and evolutionary biology will come to overlap broadly.

An interspecific comparative economics is also a possibility. During recent research on the evolution of division of labor in insect societies, George Oster and I have written a short book that resembles a textbook in microeconomics.[32] Insect economics differs in several respects: the transactions among colony members are almost exclusively instead of merely partially instinctive (that is, "irrational"), the societies are mostly sterile and female, and (because of the haplodiploid genetic bias) unilateral altruism has far greater genetic utility than in human societies. The broad forms of the analyses are otherwise much the same. The point is that human economics is not really general economics, but rather the description of economic behavior in one mammalian species with a limited range of the biological state variables.

Sociology

By virtue of its loftier perch in the hierarchy of subject matter, sociology should be the queen of the social sciences. Yet I personally find it the most alien and least interesting. Part of the cause is revealed in the following statement by Durkheim: "In a word, there is between psychology and sociology the same break in continuity as between biology and the physicochemical sciences. Consequently, every time that a social phenomenon is directly explained by a psychological phenomenon, we may be sure that the explanation is false."[33] I suspect that this statement is as completely wrong for sociology as it proved to

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

be for biology. Although few contemporary sociologists would uncritically accept *The Rules of Sociological Method*, Durkheim's dualism lives on by tradition. Sociological analysis seldom utilizes the known facts of social psychology to any depth, and evolutionary biology remains all but taboo. The specters of biologism and social Darwinism are still feared, entirely without justification.[34] The situation is so extreme that I suspect that progress in the near future will be measured by the connections sociology makes with its antidisciplines. To the extent that it does not make these connections, it will remain an ad hoc, descriptive science.

Yet sociology is not destined to be cannibalized by the antidisciplines, any more than cytology was absorbed by biochemistry. The reason is that sociology is truly the subject most remote from the fundamental principles of individual behavior. Advanced literate societies, the main concern of sociology, are the most removed in character from the kinds of social and economic systems in which the genetic basis of human social behavior evolved. Having been jerrybuilt on the Pleistocene human biogram, they are the least stable, probably have the greatest discrepancies between genetic and cultural fitness, and hence are most likely to display emergent properties not predictable from a knowledge of individual psychology alone.

The Limits of Reductionism

Karl von Frisch once made a remarkable statement about his research. He said that the honeybee is like a magic well—the more you draw from it, the more there is to draw. Other students of social insects share this sense of seemingly infinite richness in the phenomena of colonial life. They have learned that a great deal of evolutionary novelty at the social level can be generated by only a small amount of genetic change at the level of the individual. A slight modification in one parameter of allometric pupal growth, for example, can produce a new array of castes; whereas an altered response to a pheromone can create a new mode of communication.

The full phenomenology of social life cannot be predicted from a knowledge of the genetic programs of the individuals alone. When the observer shifts his attention from one level of organization to the next, he expects to find obedience to all of the laws of the levels below. But upper levels of organization require specification of the arrangement of the lower units, which in turn generates richness and the basis of new and unexpected principles. The specification can be classified into three categories: combinatoric, spatial, and historical. Thus the ammonia molecule neutralizes its electric dipole moment and conserves the laws of nuclear physics by inverting the negatively charged nitrogen back and forth through the triangle of positively charged hydrogens at a frequency of 3×10^{10} per second. But this symmetry is broken in the case of sugar and other larger organic molecules, which are too large and complexly structured to invert themselves. They break but do not repeal the symmetry laws of physics.[35] This specification may not be of great interest to particle physicists, but its effects redound throughout organic chemistry and biology.

Primitive wasps, comprising early members of the order Hymenoptera, evolved the sex determination mechanism of haplodiploidy, whereby unferti-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

lized eggs yield males and fertilized eggs yield females. This mechanism may have been a specific adaptation that permits females to choose the sex of the offspring according to the nature of the separate prey items they are able to subdue. But whatever the initial cause, haplodiploidy represented a historical accident that predisposed these insects to develop advanced forms of sociality. The reason is that it causes sisters to be more closely related genetically than mothers are to daughters, and so they find genetic profit in becoming a sterile caste specialized for the rearing of sisters. As an apparent result, social life among insects is almost limited to the phylogenetically advanced hymenopterans, namely the social wasps, social bees, and ants. Furthermore, most cases of insect social life can be classified either as matriarchies, in which queens control colonies of daughters, or as sisterhoods, in which sterile daughters control the egg-laying mothers. Many other strange effects flow from this genetic asymmetry. In addition, the hymenopterous societies have proved so successful that they dominate and alter much of the terrestrial ecosystems of the Earth.[36] Who could have guessed all of this from a knowledge of haplodiploidy?

The urge to be a reductionist is an understandable human trait. Ernst Mach captured it in the following definition: "Science may be regarded as a minimal problem consisting of the completest presentment of facts with the least possible expenditure of thought."[37] This is the sentiment of a member of the antidiscipline, impatient to set aside complexity and get on with the search for more fundamental ideas. The laws of his subject are necessary to the discipline above, they challenge and force a mentally more efficient restructuring; but they are not sufficient for its purposes. Biology is the key to human nature, and social scientists cannot afford to ignore its emerging principles. But the social sciences are potentially far richer in content. Eventually they will absorb the relevant ideas of biology and go on to beggar them by comparison.

REFERENCES

[1]G. Holton, *Thematic Origins of Scientific Thought: Kepler to Einstein* (Cambridge, Mass.: Harvard University Press, 1973).

[2]A useful ontogeny of disciplines, starting with the natural history phase, has been outlined by F. S. C. Northrop, *The Logic of the Sciences and Humanities* (New York: Macmillan, 1947).

[3]V. F. Weisskopf, "Nuclear Structure and Modern Research," *Physics Today*, 20 (1967): 23–26.

[4]An admirable example of this approach to neurobiology is provided by E. R. Kandel, *Cellular Basis of Behavior* (San Francisco: Freeman, 1976).

[5]J. S. Fruton, "The Emergence of Biochemistry," *Science*, 192 (1976): 327–334. Another recent product of this interplay is molecular immunobiology. G. M. Edelman, one of the leading contributors, has noted that "the idea of specificity and molecular recognition was prior to, and was necessary but not sufficient for, the idea of clonal selection which operated at the level of cells. In turn, the particular picture of variability and constancy which emerged at the level of the molecular analysis of antibodies could be applied to refine the idea of clonal selection at those points where it seemed most dubious." See "The Problem of Molecular Recognition by Systems," in F. J. Ayala and T. Dobzhansky (eds.), *Studies in the Philosophy of Biology* (Berkeley: University of California Press, 1974), pp. 45–56.

[6]This imbalance in experimental population genetics is well illustrated in the work of T. Dobzhansky and his associates; see T. Dobzhansky, *Genetics of the Evolutionary Process* (New York: Columbia University Press, 1970). They made brilliant progress in elucidating the genetic processes of natural selection and speciation but remained largely unaware of the central issues of ecology. In recent years some *Drosophila* geneticists have concentrated more deliberately on natural history, and information more useful for ecological analysis is now accumulating rapidly.

[7]See E. O. Wilson, "Eminent Ecologist, 1973: Robert Helmer MacArthur," *Bulletin of the Ecological Society of America*, 54 (1973): 11–12.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

[8]R. H. MacArthur, "On the Relative Abundance of Bird Species," *Proceedings of the National Academy of Sciences, United States*, 43 (1957): 293–295.

[9]See, for example, R. M. May (ed.), *Theoretical Ecology: Principles and Applications* (Philadelphia: Saunders, 1976).

[10]E. O. Wilson, *Sociobiology: The New Synthesis* (Cambridge, Mass.: Belknap Press of Harvard University Press, 1975).

[11]Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974).

[12]J. A. R. A. M. van Hooff, "A Comparat e Approach to the Phylogeny of Laughter and Smiling," in R. A. Hinde (ed.), *Non-verbal Com nication* (Cambridge: University Press, 1972), pp. 209–241.

[13]G. E. King, "Socioterritorial Units and erspecific Competition: Modern Carnivores and Early Hominids," *Journal of Anthropological Research* (1977), in press.

[14]E. O. Wilson, *Sociobiology*.

[15]D. G. Freedman has provided evidence of average differences in motor activity and excitability in neonates between Navajo, Japanese, and Caucasian populations. See *Human Infancy: An Evolutionary Perspective* (New York: Wiley, 1974). Such differences could lead to at least slight variations in the ways infants are handled and carried.

[16]L. Ehrman and P. A. Parsons, *The Genetics of Behavior* (Sunderland, Mass.: Sinauer, 1976); G. E. McClearn and J. C. DeFries, *Introduction to Behavioral Genetics* (San Francisco: Freeman, 1973); L. L. Heston and J. Shields, "Homosexuality in Twins: A Family Study and a Registry Study," *Archives of General Psychiatry*, 18 (1968): 149–160; E. H. Lenneberg, "On Explaining Language," *Science*, 164 (1969): 635–643; B. D. Davis and Patricia Flaherty (eds.), *Human Diversity: Its Causes and Social Significance* (Cambridge, Mass.: Ballinger, 1976).

[17]P. J. Richerson and R. Boyd, "A Dual Inheritance Model of the Human Evolutionary Process," *Journal of Theoretical Biology* (1977), in press.

[18]See M. W. Feldman and L. L. Cavalli-Sforza, "Cultural and Biological Evolutionary Processes: Selection for a Trait under Complex Transmission," *Theoretical Population Biology*, 9 (1976): 238–259. These authors provide an explicit model for the selection of genes that alter the learning rules of cultural transmission.

[19]M. Sahlins, *The Use and Abuse of Biology* (Ann Arbor: University of Michigan Press, 1976).

[20]Richerson and Boyd, "A Dual Inheritance Model of the Human Evolutionary Process."

[21]See for example N. A. Chagnon, "Fission in an Amazonian Tribe," *The Sciences*, 16 (1976): 14–18; and W. H. Durham, "Resource Competition and Human Aggression, Part I: A Review of Primitive War," *Quarterly Review of Biology*, 51 (1976): 385–415.

[22]E. O. Wilson, "Some Central Problems of Sociobiology," *Social Sciences Information, Biology and Social Life*, 14 (1975): 5–18.

[23]B. F. Skinner, *The Behavior of Organisms: An Experimental Analysis* (New York: Appleton, 1938). Skinner has since broadened his view to incorporate genetically programmed behavior patterns as the analogs of learned behavior in lower animals; see "The Phylogeny and Ontogeny of Behavior," *Science*, 153 (1966): 1205–1213. He differs from most biologists in his view not of the basic mechanisms of learning and instinct, but in the amount of structuring that occurs in the human brain. Matters of degree are generally the most easily resolvable in science.

[24]Summarized in M. E. P. Seligman and J. L. Hager (eds.), *Biological Boundaries of Learning* (Englewood Cliffs, N.J.: Prentice-Hall, 1972); and R. A. Hinde and J. Stevenson-Hinde (eds.), *Constraints on Learning: Limitations and Predispositions* (New York: Academic Press, 1973). A contrary view has been expressed by M. E. Bitterman, who concludes that underlying principles can still be drawn, at least for the vertebrates (*Science*, 188 (1975): 699–709).

[25]S. J. Shettleworth, "Conditioning of Domestic Chicks to Visual and Auditory Stimuli," in Seligman and Hager (eds.), *Biological Boundaries of Learning*, pp. 228–236.

[26]M. E. P. Seligman, "Phobias and Preparedness," in Seligman and Hager (eds.), *Biological Boundaries of Learning*, pp. 451–462.

[27]R. Brown, *A First Language* (Cambridge, Mass.: Harvard University Press, 1973).

[28]For example, R. J. Lifton, *The Life of the Self: Toward a New Psychology* (New York: Simon and Schuster, 1976).

[29]The same is true of Marxism, which is to a considerable degree sociobiology without biology.

[30]P. A. Samuelson, *Economics*, 10th ed. (New York: McGraw-Hill, 1976).

[31]H. Leibenstein, *Beyond Economic Man: A New Foundation for Economics* (Cambridge, Mass.: Harvard University Press, 1976).

[32]G. F. Oster and E. O. Wilson, *Caste and Ecology in the Social Insects*, in preparation.

[33]E. Durkheim, *The Rules of Sociological Method*, 8th ed., translated by S. A. Solovay and J. H. Mueller (New York: Free Press, 1938).

[34]The last remnants of social Darwinism died with the advent of sociobiology, which delineated the roles of cooperation and altruism in societies and rendered them consistent with population genetics. See, for example, the exposition by R. Dawkins, *The Selfish Gene* (New York: Oxford

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

University Press, 1976). A frankly biological critique of sociology has been provided by P. L. van den Berghe in *Man in Society: A Biosocial View* (New York: Elsevier, 1975).

[35]P. W. Anderson, "More is Different," *Science*, 177 (1972): 393–396.

[36]E. O. Wilson, *The Insect Societies* (Cambridge, Mass.: Belknap Press of Harvard University Press, 1971).

[37]E. Mach, *The Science of Mechanics*, 9th ed. (LaSalle, Ill.: Open Court, 1942).

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

EMMANUEL B. LE ROY LADURIE

Recent Historical "Discoveries"

Do we really have the right to speak of historical "discoveries"? In a sense, no. The most striking "discoveries" in this field are often the shortest lived, for example "historical demography" with its method of "reconstructing families." In more or less two decades, this discipline, initially quite fertile, has reached the stage of diminishing returns. Each new village monograph, based on a scrupulous reconstruction of hundreds of old families that had lived in the same village, adds only a few scraps of unpublished truths to the monographs that have preceded it. At this point, only the reconstruction of families on a national scale might produce something truly new.

Similarly, deductive rigor, the pride of the exact sciences, is not always so easy to attain, or so profitable, in the historical disciplines. Robert Fogel's fine and vast undertaking, concerning the economics of slavery, remains an intellectual event of primary magnitude and capital importance. But is not this very excess of mathematical rigor and vigor that surrounds *Time on the Cross*, sometimes in the absence of a totally adequate content, one of the causes, if not one of the justifications, of the strenuous debate that has followed its publication? And we wonder whether a solid quantitative work, as indeed this was, with a greater margin and greater flow of pure scholarship—at once dubitative, interrogative, qualitative, and traditionally "literary" (even in the pejorative sense of this term)—would not have had a better chance of blunting, if not indeed quite flattening out, the points of some of the criticisms? I shall limit myself to asking the question.

Unlike the exact sciences and other human sciences, history rests at the very center of an extremely crowded, multiple crossing. In certain cases it can reattach itself to the natural sciences, or rather to a "cosmological history of Nature," as would be the case, for instance, with the study of climatic evolution during the historical period of the last one or two millennia. "True scientists" will be quite surprised at seeing historians incorporate, at least in part, this territory. But their surprise will baffle only themselves. For the scientists should remember that (with the exception of dendroclimatology) the condition sine qua non for establishing a clear serial (monthly, seasonal, yearly, decennial) chronology of the European climate in the fourteenth century is to be able to read, as did the English historian Titow, the Latin scrawls of the archives of the fourteenth century. "To date scrupulously is the duty of the historian."[1]

The case of climatic history, around which various works of research have multiplied in the last years, is obviously special. But history is also connected more generally, more largely to the quasi-totality of the social sciences. At the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

head of this, as it relates to Clio, is the fundamental quadriga: sociology, anthropology, economics, demography. Such a connection appears in the case of the excellent, though sometimes questionable, *Village immobile* by the Quebec historian Gérard Bouchard. This *sologhot* village was exposed by Bouchard, who is himself bicultural, to the crossfire of French historical demography and North American sociology, both disciplines being conjugated and set in the past tense. Similarly, the "new economic history" is only econometry applied to the past, "restrospective"; and there are many more examples.

In certain cases, this "conjugation in the past tense" is quite difficult; and this is not necessarily history's fault. The responsibility should rather fall upon whatever particular social science which shows itself incapable of an "interface" with us. Let us say that the specific performances of this particular social science are, in this case, too precarious or too unfit to be the object of a projection toward the past. The best example, or at least the most topical, is that of Freudian, and even better, non-Freudian psychology. Of course, psychoanalytical history is always about to take off; it is, and will be for a long time, the *next assignment*, tomorrow morning's task, tomorrow morning being every day postponed to twenty-four hours later. If I say this, it is not to underrate the value of some great works of psychohistory already existing (Erikson, Besançon) but rather to underline the fact that this psychohistory is still very far from having reached the maturity in which bask the historical demography of Louis Henry and Pierre Goubert; the retrospective economics of Pierre Vilar and David Landes; not to speak of the sociology and the historical ethnography of Charles Tilly and Natalie Davis. This lack of maturity in relation to history is even more evident when psychology moves away from the psychoanalytical realm. Recent "social psychology," for instance, has not yet supplied any element that could be of some profit to historians. And yet these are, like birds of prey, constantly on the lookout for the new available and adaptable idea that may issue from neighboring disciplines.

Similarly, as regards linguistics, contacts with history as such have not always been very fruitful. (Of course, a good part of linguistics is historical in itself, concerned with dead languages, etc., but this is another story). Régine Robin, whose great works on the social life of France in 1789 are quite well known, has decided, with much competence, to marry the two disciplines: linguistics and history. Thus, she has cautiously started by tackling relatively easy issues such as trying to know, through linguistic analysis of speeches and presidential statements, whether, around 1960, General de Gaulle really meant to fool the "pieds noirs" (the Algerian French). Or, to determine whether *l'Union des Etudiants communistes* (subordinate to the French Communist party) was intentionally nearer to the "working class" than are the Trotskyists. In both cases, the answer, supported by linguistic sophistication, has obviously been *yes*! Which is exactly what was to be proved. But, one cannot help wondering, if this is really a discovery or, rather, a door already open burst into with a great racket.

But let us go on. On the whole, the outlook remains positive. By turning the social sciences toward the past, history has gradually advanced along the forefront, keeping its legitimacy while considerably broadening its field of action. This regressive operation, starting from contemporary or recent social thought,

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

has become all the easier and more fruitful when it was dealing with the great thinkers of the nineteenth and early twentieth centuries, whose thought is, of course, very near to that of professional historians. I am thinking of Marx, Weber, Freud, Durkheim. A book like Mona Ozouf's on *la Fête révolutionnaire*, for instance, is a rigorous example of the application of the Durkheimian method to the large dossier on political and ceremonial festivities at the end of the eighteenth century. In this particular realm, Ouzuf has not made any "discovery" as such but she has added the relevance of supplementary, historical past to the initial Durkheimian insight as relative to the integrating functions of holidays. And this alone is quite an achievement. Moreover, in this very delicate business concerning the extension of interdisciplinary fields, history has been protected by its label. To get a hold on the text of Marx and Freud in order to apply, with good reason, their intuitions to the past, does not mean that we force Marx-and-Freud (always strictly linked) into saying whatever accords with the straight or tortuous logic of this or that temporary current of Parisian fashion. From this point of view, history is, thank God, solidly protected against such procedures, as if, given its bulk of traditional methods, all firmly established from the times of Herodotus and Thucydides, it were immune to them. History has been able to defend itself against such temporary temptations much better than, in Latin countries, some very "engaged" segments of the social sciences. In this last case, Leftist temptations have led to works, not to mention university courses, that are biased and repetitive. In the historical realm, damages of this kind are limited, perpetrated as they are by at most one or two works[2] which can easily be viewed as the exceptions that confirm the rule.

I now come to another reason which makes it very difficult to apply the term "discovery" to the investigations of historians. The discipline that they practice, bloated as it is, does not limit itself simply to absorb a small part of the natural sciences, and the greatest part of the social sciences applied to the past. History also insists in maintaining strong ties with purely humanistic disciplines such as literature and art. In fact, it is even in close contact with some forms of show business, such as theater, opera, cinema, and also—horrors—television! I beg my readers not to throw up their hands in despair when they read this sentence, and particularly the term that concludes it, television. I beseech them not to forget that history still depends on imagination, fantasy, and sensitivity, as much as on intelligence and collective memory. In fact, history cannot justify its existence by the extraordinary rigor of the results, however disinterested they may be, that astrophysicists are used to obtaining in their own realm. It cannot even pride itself with services rendered to the human community, as is the case with, for instance, Pasteurian microbiology, or with genetics at the time of the "green revolution." Nor can it put itself to the practical disposition of the administration, as is currently done by sociologists when, on behalf of state planners, they determine the possibilities for touristic development in a particular region, or the popular hostility that will be aroused by the local establishment of a given industry. The last time that French and German historians entered the service of their countries was at the eve of World War I in order to justify the mutual preparation of the great massacre. No wonder that, given the outcome of that drama, our profession is so reluctant to repeat the experience.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

In fact, history answers both an individual and a collective need for serious thought; it answers a curiosity, a desire; it answers the search for pleasure in reading (for the reader), and often for pleasure in writing (for the writer of history). In this sense, the décor and the display are as important as the merchandise, and the style as important as the thought, from which it is indeed inseparable, and whose validity it upholds with its pertinence and its impact. The current success of the works by Fernand Braudel, at once literary and scientific, speculative and aesthetic, cannot be explained otherwise. It is these very same reasons that, in the realm of the social sciences, fully justify the success of Barthes' work, which, by means of his own unique skill and style, was able to turn an otherwise unreadable semiology into a pleasant and readable work (see Barthes' fine book on *S/Z*). It is easy to understand how in such a literary ambience, "invention" in the most purely scientific sense of the term, may have problems finding its right place.

Does this mean that the concept of "discovery" is completely inoperative or inadequate as respects history? This is not the case, especially if we are allowed to modify, however slightly, the meaning of the term and can give it the same meaning that miners give it when they speak of a "discovered" coal mine (that has been brought into the open by means of the removal of the most superficial layer of soil). In some ways, and without trying to be paradoxical, we could even go as far as saying that the historian of 1977 is better placed to make such discoveries than the physicists or the chemists of our time, and certainly at a lesser cost. And surely these two disciplines, physics and chemistry, for different reasons, often give the uninitiated the no doubt superficial and erroneous impression of being "at the end of their tethers." Both are, of course, perfectly capable of discoveries, even fundamental ones, but often at the cost of millions and millions of dollars. And, in order to quote an example from yet another discipline, let us say that today in general we are unfortunately no longer in Galileo's time, when a man of genius could, by means of a relatively simple and inexpensive telescope, make all kinds of essential finds in various realms of astronomy. Curiously enough, as regards history proper we have not yet come to the twentieth century of ruinous particle accelerators. We are still in the inexpensive seventeenth century of Galileo, or in Darwin's nineteenth. This does not mean that I consider current historians as so many Galileos or Darwins, not in the least. That would be ludicrous. What I mean is that this very state of underdevelopment of our discipline, and the quantity of archives still quite untouched or barely touched, allow the historian, whether financially secure or not, to undertake, in the vast unexplored realms that surround his knowledge, several incursions and forays that will cost him relatively little both in money and individual time. These incursions are much more similar to the practice of the natural sciences 100, 200, or even 300 years ago than to what this has become in our time, in this glorious present where in physics powerhammers are often necessary to crush a fly, but then again, what a precious fly!

From this vantage point in History, the French system of the Major Thesis, as archaic and compilatory as it may seem at first, in the end has not proved completely negative at all. A young scholar is unleashed in the archives of a department or a province for ten, or even fifteen years. Obviously he grows a little less young, and his hair a little more gray, and his advancement is unfortu-

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

nately, at least according to the American standards, outrageously inadequate. But it is an ill wind indeed that blows no good at all. For in certain cases this method allows the researcher to accumulate real treasures of information, by far more substantial than if he had finished his Ph.D. in twenty-four months. Obviously the favorable results obtained in France would not have been possible if, in 1912, Lucien Febvre, with his great thesis on the Franche-Comté of Philip II, had not turned the massive provincial monograph on the history of the urban and rural masses into a historiographical genre as respectable as was classical tragedy in the literature of the seventeenth century. More generally, the enormous variety or freedom of human behavior gives the historian who observes it an option to invention (in the pedestrian literary sense of the word) that is quite forbidden to the physicist or the biologist: since these two individuals are professionally involved in processes that are clearly more repetitive and more monotonous than is the "human phenomenon" in itself.

And now let us touch upon a few examples relative if not to the discoveries (given its ambiguity, I prefer not to use this term) at least to the "finds" of historians in roughly these last ten years. Under the circumstances, my point of view will be dominated (though not exclusively, I dare hope) by a certain French provincialism. My English-speaking readers will forgive me for such an exclusive form of parochialism; but in this manner I may provide them with information that is sometimes completely new to them. Maybe someday they will reciprocate.

"Social history" has been holding a place of honor these last years. I take the word "social" as meaning "sociological" of course: in the first place the historical sociology of specific groups. I am thinking, for instance, of the enormous thesis by Jean Mayer, a true chunk of Armorican granite (reduced to one-third its size in the paperback edition), concerning Breton nobility in the modern age, from the seventeenth century to the nineteenth, with particular emphasis on the eighteenth. This book touches upon several fascinating points about the history of inequality, first, the birth of a caste, or rather its definitive or temporary closure. Indeed, we know that the minister Colbert had demanded, in his great reforms, the "filing," or the registration in writing, of all the nobles, to avoid the proliferation of false aristocrats, who would otherwise be granted illegitimate fiscal exemptions. But, as is often the case, an administrative measure can have consequences that its author had not foreseen. By the last third of the seventeenth century, the Colbertian reforms had transformed the community of Breton nobles into a real caste, almost as closed as could be a Brahmin family group. Now, paradoxically (at least for all those who cannot understand the permanence and the resilience of *Homo hierarchicus*) this caste, far from declining, would later expand in the material and cultural enrichment of the eighteenth century. And what is more, far from crumbling during the following stage between 1789 and 1799, it would move beyond this tragic decade and experience a second youth, thanks to the support that Breton peasants, now royalists, would grant it. This peasant mutation, so favorable to the interests of the peninsular aristocracy, was due to the villagers' disgust for the republican bourgeoisie of the Breton towns, which they had perhaps considered as an ally in 1789, but which they had fast come to consider as their main exploiter.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Other avatars of a specific group, but whose reach is this time national, and even international: the ones that are concerned with the French army, in its medieval version, as studied by Philippe Contamine. With our warmest approval, this learned medievalist scholar dismantles the Procrustian frame of traditional and bellicose military history at all levels. Contamine also considers himself as the ethnographer of a communal soldiery, and the anthropologist of a global society (France), as judged through a profession (the military one). It is after 1450 that a permanent, regularly salaried army of 10,000 men makes its definitive appearance; a "strike force" that no one dares question any longer, even though it costs the country a great deal. The professional soldier, this currently threatening character, is hence one of the inventions of Renaissance Europe, as is the compass, America, and the printing press. With the invention of these military men, France also invented the civilian—two birds with one stone. The paradoxical result was the demilitarization of the global society, and not the opposite. All these intuitions, strongly supported by extensive documentation, turn Contamine's fine book into one of the keys of our modernity, even though this rigorous work appears singularly devoid, we are sad to say, of military truculence and coarseness.

Whether it be about Breton nobility, or the French army, the monograph is, of course, not an end in itself. With most avant-garde historians, the study of this or that social group always ends up as a theory of elites; a theory that, since Pareto, has been quite familiar to sociological thought. Denis Richet has introduced the notion of "elites" in the French historiography of our "Enlightenment." This notion has, furthermore, the immense advantage of transcending the opposition that "Marxism" (somewhat simplified, no doubt) has carved out between the bourgeoisie (revolutionary) and the nobility (ipse facto reactionary). In fact, in the age of Louis XV and Louis XVI, the elites regroup the enlightened portions of the nobility (magistrates, etc.), the financiers (whose nobility is recent, but bespattered with wealth); a part of the high clergy that flirts with philosophy (canons, vicars-general); and the fashionable segment of the bourgeoisie (lawyers, intellectuals, etc.). As for the commercial and capitalistic-industrial bourgeoisie, it would seem to hold, if we are to believe certain glosses and certain gnosses, the keys to a radiant future. Unfortunately, "nasty facts deny beautiful theories." This fraction of the French bourgeoisie stands, in fact, apart from the enlightened and progressive movement that remains inserted in this single yet composite elite of Lights of the Salons and the Academies, which I have tried to describe below with the help of the works by Richet, Furet, Roche, and others—an elite that digs the grave of the Old Regime, and sometimes even of its own. These last remarks are particularly true of the enlightened and liberal nobility, an attractive estate that is as clairvoyant as respects the interests of the nation as it is blind to its own. Altogether, these different "peak" groups, without knowing it or suspecting it, clear the ground for the French Revolution.

This rejuvenated conception of the elite, as proposed in the last ten to fifteen years, will hence renew, for instance, the entire historical theory concerned with nobility. The Breton case, briefly mentioned above, is marginal. It would therefore be better to widen the scope to the national scale. The French nobility, if we are to believe one of its recent historians, Guy Chaussinand-Nogaret,

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

is not essentially the privileged, obtuse class that wanted to reverse the march of history, depicted in vivid colors by the neo-Babouvian historiography of the French Revolution. In fact, it is a question, in a completely different perspective, of a numerically small group to be sure (less than 200,000 people), but also one that was economically dynamic and even capitalistic, in the best sense of the term; an intellectually liberal group, and eventually a very advanced one. The Cahiers of the Estates General drawn up by the nobility in 1789 are as revolutionary as those of the Third Estate. Later on, of course, the great majority of the nobility would turn very rapidly against the French Revolution once this was under way. But this change of attitude came late. It should not falsely color, by retrospective teleology, our entire vision of the mobiliary worlds of the Old Regime in the eighteenth century. What would we say of a history of France that elaborated its entire vision of the twentieth century according to an unfortunate social revolution that had happened rather late, in 1978, for instance?

But let us remain before 1800, or before 1815. All this "elitist" sociological revisionism has clearly affected or interpretations of the French Revolution. In order not to fall into platitude, these must by necessity differ from the essentially bourgeois and "Third Estate" version that was given by l'abbé Sieyès himself ("What is the Third Estate? Everything. What has it been until now in the political order? Nothing. What does it want? To become something.") More prevalent than ever in our day is the idea (already superbly formulated by Georges Lefebvre) that there is not one single bourgeois, objectively procapitalistic revolution that forms a totalizing block, but rather several revolutions: one certainly bourgeois, but the other specifically peasant. Intelligent *jacquerie*. It furthers the familial, "lopinist" exploitation of the small farmer against the large manorial and domanial economy, which has an already pronounced vocation for capitalism. There is also the plebeian *sans-culotte* revolution of the cities, with its simultaneously *passéiste* and futurist tendency. And then the teratological Terror of 1793–1794, already prophetically announcing, as the spermatozoon already prefigures the man, certain totalitarianisms of the twentieth century. In this sense the works of François Furet are paradoxically similar to those of Richard Cobb, despite the polemics launched by the English historian against his French colleague. Hostile to the elementary platitudes that show us the "bourgeoisie" allied to the "people" against "feudalism," in effect Cobb insists on different sacreligious or iconoclastic themes. He underlines the opposition between starving city and bountiful country during the French Revolution. He emphasizes the universality of the popular movement, which involves the "white" terrorist as well as the "red" or "blue" *sans-culotte*. Finally, speaking of the *sans-culottes*, Cobb justly stresses the marginal role of their action. An excited troublemaker, the popular militant, according to Cobb—to whom I leave the responsibility of his assertion—is an "aberration of Nature." He stirs up the masses, but does not really represent them. He speaks for them, without having their mandate. It is the "eternal problem" of the active minorities: they will be frequently busy, for better and often for worse, in the twentieth century.

Historians come across these active minorities again, ages ago, well before the French Revolution, during the centuries of the Old Regime. This involves the major question of the popular revolts of the seventeenth century, which let

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

rivers of blood flow under Louis XVII, and streams of ink after 1960: in effect, Porchnev and Mousnier have given a nonnegligeable importance to these uprisings in a controversy that resounded throughout the small world of historians. After these quarrels of Titans, which had the great merit of getting the ball rolling and initiating debate, subsequent researchers, more modest but just as effective, set themselves to the task. Yves Bercé, a student of Mousnier, published a large look in which he gave a typology of the popular revolts in Southwestern France between 1590 and 1710. According to him, these revolts do not represent, everything considered, a disorderly explosion, but rather a regularly recurrent pattern of cultural behavior. They are not directed solely, or even essentially, against manorial nobility, as implied by the *images d'Epinal* pertaining to the model Jacquerie, in which the country folk slaughtered the nobles and sodomized their ladies. Rather they are especially directed against the State, against excessive or unjust taxation, against the billeting of soldiers, the price of bread. They are rational, just as strikes are in our own days (at least in the rational framework of the strikers, which does not necessarily coincide with that of the national economy and the fight against inflation). These old revolts are rational because they often succeed in reaching their stated objectives, even though their leaders may be hanged or beheaded in the process. (To have a comparative idea of this curious phenomenon in our days, we should try to imagine a strike of the workers of the Ford factories in Detroit: chances are that they would obtain their raises, and that the leader of their union would end up in the electric chair). The old-fashioned revolt is not only planned according to an often valid end, but it is also emotional; for, in the celebration and the excitement of the barricades, it often manages to realize the carnival of a topsy-turvy world.

All these uprisings are carrying us, or so it seems, to the antipodes of the question concerning the elites. But, in fact, we have remained within the margins! After the higher margins (the leading classes), we have shifted to the lower margins (inferior groups, and even "proletarians in rags": the *lumpenproletariat*). In this context, the poor of yore have become one of the most important subjects for historians, just as, today, they are one of the most sensitive points of the urban crisis of the twentieth century, although the poor of the preindustrial society had the same skin color as the rich. The Middle Ages, explored by Michel Mollat and his students, is obviously one of the most privileged periods for the study of overwhelming pauperism. Who are the poor of this period, according to the works of Mollat's team? Well, in the first place the wage-earners. In the Florence of the fourteenth century (and not only there!) they are almost always victims of a certain penury, from the very moment they get the "troublesome" idea of getting married and hence of having children; in other words, from the moment the good life of their bachelorhood ceases—a "good life" at the gastric level, but not always at a sexual, or familial level. Wage-earners aside, the poor are also those who are recognized as such by their social status, the lowest on the scale of disdain. By virtue of this they are given alms by the rich, who are their counterparts at the other end of the social ladder. These "true" poor are, quite simply, bums, beggars, the poorest of the poor, the lowest of the low. They are the image of Christ on earth, which is why they receive alms advantageous to the heavenly salvation of the giver. Then there are

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

the adherents of voluntary poverty, the formerly wealthy young men who impoverished themselves in order to imitate Christ, in other words, the hippies of the period à la Saint Francis of Assisi. All this—with the exception of the Franciscans!—will flow into the great confinement of the poor and the marginal in the modern age evoked by Michel Foucault's well-known book.

All these various problems or topics—social groups, elites, revolution, the poor—border on Marxism without ever confronting it face to face. Up to this point we have been aware of only the profile of this great ideological system. Should one then take the bull by the horns? It is a fact that, in all the studies of social history, the researcher utilizes ipso facto certain Marxist (and often pre-Marxist) concepts such as social class, class struggle, and even mode of production. But one should avoid like the plague all the dogmatic, rigid, absurd elements present in the now standard version (at least as concerns the Latin countries) of Marxist-Leninism, or of crypto-Stalinism. And we cannot help wondering whether a history that claims to be completely non-Marxist would not have a certain interest and certain advantages! Of course, as I say this, I feel as if I were walking on eggs, as if I had profaned the most sacred vessels of our times. Never mind. A history that is completely devoid of all the trivialities of Marxism would be of great value for these countries whose social structures are allegedly modeled on that doctrine, and in which that doctrine has in fact become its very opposite. Is this not why, in a certain sense, Solzhenitsyn has been able to write an exemplary history of some aspect of past life in the Soviet Union, and more specifically of Gulag? In order to attain this objective, he had entirely ignored all the ideological postulates officially, though hypocritically, espoused by Soviet society. Likewise, Philippe Robrieux in writing his biography of Therez has not used the Marxist approach, irrelevant in this case, but rather the good old methods proper to *la petite histoire* and the supermodern ethnography of oral surveys. And the young historian Emmanuel Todd, in his challenging essay on the Soviet Union, has successfully approached current Communist society with certain techniques of enquiry that had been previously employed by Pierre Goubert, in his study of the Beauvaisis in the seventeenth century.

"In my Father's house are many mansions," and there are also a variety of possible points of departure in the practice of social history. One of these points is the one that has situated the minorities (which history has silenced without their being destroyed) at one of its extremes, and at the other the majorities (silent); this expression "silent majority" has, I hasten to say, a meaning that is quite different from the one that is commonly applied to it today by the orthodox Right.

First of all, the minorities. The sixties have brought back into fashion the "small" peripheral nationalities of the "great" European nations: Basques and Catalans, both astride the Franco-Spanish Frontier of the Pyrenees; Occitans and Bretons in France; Scotsmen in Great Britain; French-speaking Jurassians from the Swiss canton of Berne (all proportions considered), etc. Historians, and why not, have fallen into step with this fashion. André Burguière and Suzanne Berger, for instance, in their respective works *Bretons de Plozevet* and *Les Paysans contre la Politique*, have contributed two remarkable studies about

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Brittany. These studies show quite clearly how in this small peninsula, just as in Roman Catholic Ireland and Roman Catholic Poland, church organisms during the nineteenth and twentieth centuries played a very positive role in the establishment of a sociability (in this particular case, agricultural syndicalism) that defends the local structures while pushing them to evolve in the right direction: the results, if not the aims, of this action are the opposite of narrow clericalism. The traditionalism of the minorities can hence be at once a factor of both conservatism and progress (since these two terms, as we know, are neither contradictory nor mutually exclusive). These minorities are also evidence of the tremendous time-rooted nature of old conflicts: thus we have found out that, in Roman Catholic Bretagne, the "red" inclinations of the village of Plozevet, quite exceptional in this duchy, date back to 1789; they have persisted, under various forms, to our day.

And now the majorities. In France, the true, the great "silent majority" of the Old Regime, and by this I mean the peasants (who constituted more than 80 percent of the population during the eighteenth century) has received its share, and more than its share, of historical consideration. The "peasant tradition" in French historiography can be dated back, for instance, to Marc Bloch (*Caractères Originaux*) or to Georges Lefebvre (*Paysans du Nord*). Georges Lefebvre has shown how the small peasantry of the département du Nord, familial, divided into small portions, and superproductive in Flemish or Chinese fashion, had been able to lay hold of the manorial and clerical lands by means of the French Revolution and of the sales of national property.

In a recent book, Jean Jacquart comes up with the exactly opposite scheme. But it is also true that he deals with a different region (the country around Paris) and a different period (the sixteenth and seventeenth centuries). Thus, this second model does not invalidate the first: it simply shows the other side of the alternative. Jacquart does not celebrate the triumph of small familial agriculture, but rather its defeat by the offensive of the "large enterprises" teleguided by the urban nobility and the Parisian *notables* of the classical age. With the strength of their capital, these groups install large farms and powerful capitalist farmers all around Paris for which their large estates constitute a stable source of income and wheat. Thus we have a good outline of the two essential vectors of the evolution of our countryside under the Old Regime: the familial-parcelled-out-revolutionary vector (G. Lefebvre); the manorial-big-farming-oriented-capitalist vector (J. Jacquart).

But history has also galloped, at the risk of losing its breath, on the new hunting grounds of the votaries of Clio, who have now turned into sexologists, gastrologists, ethnologists. A sexual historian, and professor at the University of Grenoble, Robert Solé has described, in a sometimes smutty style wallowing amidst breasts, genitals, and buttocks, the orgies and other baroquisms of the flesh of which the men and women of the Renaissance and even the classical age were "guilty" or responsible. Who kisses too much embraces poorly? Maybe so. In any case, Solé's vision seems to have been too broad; this compendium of all the sexual deviations and superlative normalities far exceeded the dimensions of a single book. But the Grenoble historian has shown a lighter touch when dealing with the great antimasturbatory offensive of the eighteenth and nineteenth

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

centuries. Whether it be pre- and postrevolutionary France, or China after 1960, it would seem that any society that initially involves itself in a determined program or process of primitive birth control by such means as coitus interruptus, must pass through a phase of sex education which involves—God only knows why but it is a fact—a firm repression of masturbation.

More prudent and level-headed than Solé, Jean-Louis Flandrin has been wise enough not to separate the history of sexuality from that, finally more dominant, of the family. On this point he has written several substantial, though often complicated, books. He has demonstrated, maybe in opposition to what Ed. Shorter had somewhat hastily shown in a brilliant essay, that romantic love did exist, however corseted by social structures, among the peasants of the Old Regime. Flandrin has also tried to answer the irritating question asked by several historians: why has France, less developed economically, preceded England (economically more advanced) on the road to a contraceptive revolution since the decade 1780,[3] and particularly since the decade of 1790 and the French Revolution? In a very intelligent, though not always very convincing, demonstration, Flandrin attributes this French precession or, one could say, "priority" to various factors. Prominent among these, and very largely diffused among the elites, are the models of pro-feminine gallantry and protection of childhood spread by the Catholic Counter-Reformation and the *Précieuses* and *Salonnardes* of the seventeenth and eighteenth centuries. These elitist models seem to have had a delayed impact on large masses of the population, at the end of the eighteenth century, by means of a process of cultural diffusion and, eventually, secularization. All this is, at once, attractive and fragile; culturalist, problematic—and maybe undemonstrable. As far as all this is concerned we remain, and I beg the reader to realize that this is not meant to be a criticism of Flandrin, at the level of the most shimmering "literary" approximation, rather than at that of scientific rigor, both demonstrative and correlative. Which brings us back to certain considerations that were raised at the beginning of this article.

Once, people were used to starving, or, at least, to being quite hungry. More than they are today. Alimentary problems were far more acute in the past than in our days, and were especially characteristic of these countries which today are the most highly developed. It is therefore normal that concern for the stomach was then, with all due deference to Freud, far more intense and important than any questions pertaining to sex, however lancinating these may seem in 1977. From this can be derived whole series of consequences at the level of traditional mentalities, even among well-fed elites. Jean-Paul Aron, for instance, has shown how the French bourgeoisie of the nineteenth century had staked its honor on the "grande bouffe," the wines and the roasts, just as once Chivalry has staked its own on the quest of the Holy Grail. Norbert Elias, instead, has limited himself to attack manners: why have farting, belching, and even spitting, so anodine in itself, become clandestine, why have they gone underground? Why has medieval courtliness, so physically promiscuous (people were used to eating out of the same pots, and to wiping their mouths with the same napkin), turned into our modern and slightly icy etiquette, where physical contact is avoided as much as possible among well-brought-up people, and where each of us goes about in his invisible, portable ivory tower?

Let us return, for a moment, to the question of sex, so dear to Flandrin and

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Solé. In the general realm of manners, this particular sector is indeed best studied by historians, at least at the level of a quantification that, of course, remains quite rough. As a matter of fact it profits from the immense contribution of historical demography, and this despite the previously mentioned diminishing returns of this discipline. Would we have been able, even if only for a single moment, to imagine the tremendous sexual austerity of the French lower classes in the seventeenth century, for instance, if, for the last ten to fifteen years, we had not had the statistics compiled by the demographers at our disposal? These show the insignificance of prenuptial conceptions in the past, whereas today, despite our extraordinarily contraceptive age, such conceptions make up for at least one-fourth of the births. The fact that the England of the seventeenth century, I am sad to say, showed itself far less virtuous than France in this particular context, only points to a large insular exception: it confirms the continental rule of austerity which, we must admit, itself entails numerous splotches.

Thus, historical demography is not simply a tedious accumulation of "small strokes," otherwise called men and women, who go through the most basic process of being born, mating legally, and dying (counting, on the parish registers the numbers of christenings, weddings, and burials). If this discipline consisted only of such tasks, it would soon turn all those who practice it, including its most eminent priests, into some sort of Sergeant Flick, whose sole motto would be: *Count, count your men; count, count them well*. In fact, as we have seen, historical demography is also the study of the human concern with sex, the body, medicine, and death. In *Les hommes et la peste*, the physician-historian-demographer Jean-Noël Biraben describes what the black plague of the fourteenth century meant, both quantitatively and qualitatively, in terms of the unexpected or unhoped-for result of the microbic unification of the Eurasian world, a unification that was first brought about by the armies of Gengis Khan.[4] This pandemic was the equivalent of what would be, today, in terms of human losses, an atomic or bacteriological war of average size.

With J. N. Biraben, and even more so with M. Vovelle, F. Lebrun, and P. Ariès, death is placed at the center of historiography, "like the cemetery in the middle of the village." Vovelle must take the credit for having created the subject of "death," which only ten years ago would have been unthinkable in our discipline. It is he who, above all, has legitimately integrated this subject into the cultural currents of the Enlightenment and the Anti-Enlightenment in the eighteenth century, in Provence. Statistics in hand, drawn from the stereotyped but revealing invocations that are contained in thousands of wills, we can hence prove that the typical if not average Provençal dies as a baroque Catholic around 1720; as a Jansenist around 1750; and already as a quasi-Voltairian around 1780.

In relation to Vovelle's authoritative oeuvre, the works of F. Lebrun incline more toward an economics-sociology-demography of death. In a large book, Lebrun has described the Old Regime paradox of Anjou, this lovely region of the Loire Valley, with such an extreme rate of mortality: various epidemic plagues, and particularly dysentery, with spasmodic recurrence, prevented all demographic development in this region throughout the eighteenth century. And this is an extraordinary exception (which one could however find, though to a lesser extent, in other regions of Western France, such as Brittany, Normandy, etc.). It is a remarkable exception, since during the same period (the

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

eighteenth century), all the countries of the world, from France to China, and from North America to Russia, generally experience an increase in population: an increase that is without precedent since the thirteenth century.

Still within the domain of death, Ariès, as a cultural historian, has worked at a very long periodization. Death is, at first, familiar (Middle Ages), then macabre (fifteenth century), then baroque (seventeenth century and beginning of the eighteenth), then familial and intimate (end of the eighteenth and nineteenth centuries); and finally obscene (maturing twentieth century). What will be the next stage?

And yet death exists only in relation to life. The historians of annihilation do not specialize exclusively in this gloomy subject. In effect, how could they cramp themselves so much, without precociously digging their own graves? Hence Ariès, before getting interested in death, was the historian of childhood, the family, and even of Time (one of his least well-known and most remarkable investigations). Vovelle, instead, had a very close look at death before moving on to festivity. The festive phenomenon is, in fact, the supreme, paroxistic instance of the integration-disintegration-reintegration of an urban or a rural community, an instance that can be either conservative or revolutionary, depending on the case, the place, the time.

Even when festive, Vovelle's history is fundamentally quantitative, based on scrupulous compilations of various festivities, under the circumstances. From this point of view, it is quite obvious that quantitative history today has gotten the upper hand on both sides of the Atlantic. Indeed, how could we formulate any generalization if, sooner or later, this was not supported by a balance of frequencies, by a percentage, by a quantity of facts and, let us hope, by significant statistics? Beyond counting by hand, the use of the computer constitutes the subsequent, indeed prejudicial, stage. This is, of course, easier to realize on the American campus, which is well equipped from this point of view, whereas all too often the universities of Western Europe, even those which can boast of excellent historians among their faculties, are electronically poor insofar as both hardware and particularly software go. From such a point of view, the "Americanization" and the computerization of our customs, as historians, may well be desirable, at least in this particular sector. But it also presents unquestionable dangers: information processing brings a huge supplementary brain within the reach of the first imbecile (and there may be some in every profession, including ours). And in this case it would end up multiplying idiocies. And this is why we should never forget that the job of the historian, even when equipped with computers, remains a craftmanship of art, superlatively qualified. It is an "aristocratic" profession, just as cabinetwork and inlaying once were in the Faubourg Saint-Antione in 1789. It is a profession that is open to all insofar as birth is concerned, but accessible only to a small number of people insofar as talent goes.

An aristocrat of figures, fluently blending quantity and quality, under the auspices of a benevolent Marxism—such is, for instance, Pierre Vilar, as attested to by his book on the general history of coinage in the world during the last six centuries. The great merit of this work is not to study gold alone, but rather always to situate treasures in their causalizing and causalized relations

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

with the general decreases and increases of production, prices, populations. In Pierre Vilar's wake, Guy Bois has recently appeared. The Normandy of the Middle Ages and the Renaissance that he describes for us, shaded by quantified apple trees, is that of Malthus and Ricardo, both guaranteed by Lenin's bourgeois caution. A whole ecosystem (population, land distribution, price of wages, groundrent, and I omit the better ones) disintegrates through the pages of his book, from the black plague to Joan of Arc, and then rebuilds itself symmetrically from the end of the Hundred Years' War to Henry II. This second phase of long and schematic recuperation is what is commonly known, in economic and demographic terms, as the Renaissance. This bisecular balancing, according to Guy Bois, entails a certain drift toward a small-scale, rural capitalism.

As applied to urban history and to the problems connected with ascending and descending social mobility, the quantitative method, such as it is practiced in the United States, has given us a few masterpieces: I am thinking of Thernstrom's works. Even in France (though the fact is less well known), a great book has sealed the triple alliance among history, computation, and the city. In *Lyon et les lyonnais au 18e siècle*, Maurice Garden has given the demography of a very large town with its swarming dimension and its weight of human suffering.[5] Lyon is a cradle city: its average production of babies per female belly is by far higher than that of moderate rural demography. This urban supernatality derived from the fact that most babies were put out to nurse, which deprived the young mothers of that temporary period of sterility statistically induced by the act of nursing. On the other hand, these babies died in great numbers in the homes of their mercenary nurses—the very ones who had been the involuntary cause of their birth. A city of cradles, Lyon is thus also—and for the same reasons!—a city of graves. In the seventeenth and eighteenth centuries, Lyon is capable of maintaining its demographic balance only because of a constant immigration of men and women from the nearby villages. Hence, the great city of yore is not a pole of growth, but rather a safety valve for the surplus peasant population. Paradoxically, it functions as a factor of demographic reequilibration in a global society (urban + rural) that is poor and threatened by overpopulation.

Garden's *Lyon* is a monograph. French historians, one might say, are particularly good at specialized monographs, and also good at editing great, purely local texts such as *Pierre Rivière*, the crazy Norman peasant and criminal of some genius, whose fascinating autobiography Michel Foucault published not long ago.

This reproach, if it is one, as to the overspecialized character of our research, seems unfair to me. Just the name of Braudel and the world-wide scale of his *Mediterranée* is enough to refute any such accusation. However, it is true that when one wants to find a good, broadly sympathetic book on the economic history of Europe, it is sometimes better to go look for it in America. David Landes, for instance, with his *Europe technicienne*, is the deserving pioneer of a global vision, on a Western scale (England + Belgium + France + Germany), of the heroic as well as the recent times of the Industrial Revolution.

One could also reproach the best works of the new historical school for being

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

concerned only with remote centuries. The answer is easy: these works deal with the Braudelian concept of *long duration* [*longue durée*] (stabile, immobile, almost immutable). But "long duration is no longer a thing of this world."[6] Or at least not altogether. It has been factually expelled from the twentieth century by the acceleration of techniques and of "progress," by uninterrupted, kaleidoscopical social change. It is therefore normal that the masterpieces of contemporary historiography are more concerned with the seventeenth century than with the twentieth. It is deplorable, but it is a fact.

I have mentioned the word masterpiece: it evokes the supreme achievement of the candidate which qualifies him as a Master in medieval guilds. And, without having to go as far as that, it is true that historical discovery is quite different from what is called by that name in the exact sciences. It is all up to the creator: he leaves behind himself cultural monuments, elaborated out of vast and precious ensembles, the latter having been recovered from the depths of archives, after the removal of superficial layers. Twenty years ago such assertions, on my part, would have caused any physicist to smile. But perhaps the inferiority complex and certainly the feeling of guilt have occasionally changed camps in 1977. In fact, does not this guilt haunt today's nuclear physicist dismayed by certain dangers far more than it affects the human sciences? Having forsaken its nationalistic deviations of 1914, history is no longer a pollutant; it represents an activity that is relatively uncostly, and culturally fecund. The French Republic is proud of the presence, on its soil, of eminent physicists and great biologists. But this does not prevent this same Republic from continuing to produce good historians, just as it has excellent wines, renowned restaurants, and Camemberts of the finest quality.

Translated by R. Sieburth

References

[1]P. Vilar, in J. LeGoff and P. Nora, *Faire de l'histoire* (Paris, 1974), vol. 1, p. 185.

[2]J. Chesneaux, *Du passé fasions table rase* (Paris, 1975).

[3]Yves Blayo, "Mouvement naturel de la population francaise," *Population*, November 1925, p. 40.

[4]On this subject, see my articles "L'Unification microbienne du monde," *Revue Suisse d'histoire*, 1973, pp. 673–692, and "L'Histoire immobile," *Annales*, 1974. Later on, W. H. McNeill has worked on analogous ideas in his book *Plagues and People* (New York: 1976).

[5]Maurice Garden, *Lyon et les Lyonnais au 18e siècle* (Paris, 1974, 1975).

[6]P. Vilar, in *Faire de l'histoire*, vol. 1, p. 197.

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

Notes on Contributors

JONATHAN CULLER, born in 1944 in Cleveland, Ohio, is Professor of English and Comparative Literature at Cornell University. His publications include *Flaubert: The Uses of Uncertainty* (1974), *Structuralist Poetics* (1975), and *Saussure* (1976, 1977). *Structuralist Poetics* was awarded the James Russell Lowell Prize for the best work of literary criticism published in 1975.

SHMUEL N. EISENSTADT, born in 1923 in Warsaw, is Professor of Sociology at the Hebrew University, Jerusalem. Among his publications are *From Generation to Generation* (1956); *Social Differentiation and Stratification* (1971); *Tradition, Change, and Modernity* (1973); and *Revolutions and the Dynamics of Civilizations* (forthcoming).

EMMANUEL B. LE ROY LADURIE, born in 1929, is professor at the Collège de France. Among his publications are *Les Paysans de Languedoc* (1966, 1974), *Histoire du climat depuis l'an mil* (1967), *Times of Feast, Times of Famine* (1971), and *Montaillou, village occitan, 1294 à 1324* (1975).

GEORGE A. MILLER, born in 1920 in Charleston, W.Va., is professor at The Rockefeller University and has published *Language and Communication* (1951), *Psychology, the Science of Mental Life* (1973), and, with P. N. Johnson-Laird, *Language and Perception* (1976).

BRUNO ROSSI, born in 1905 in Venice, is Institute Professor, Emeritus, at the Massachusetts Institute of Technology. In addition to numerous papers, his publications include *Rayons Cosmiques* (1935), *High-Energy Particles* (1952), and, with S. Olbert, *Introduction to the Physics of Space* (1970).

JEAN STAROBINSKI, born in 1920 in Geneva, is professor at the Faculté des Lettres, University of Geneva. Among his works are *Jean-Jacques Rousseau: La Transparence et l'obstacle* (1957), *L'Oeil vivant* (1961), and *Les Emblèmes de la raison* (1974).

LESTER C. THUROW, born in 1938 in Livingston, Mont., is Professor of Economics and Management at the Massachusetts Institute of Technology. He is the author of *Poverty and Discrimination* (1969), *Generating Inequality* (1975), and "Optimum Trade Restrictions," *Econometrica* (1976).

STEVEN WEINBERG, born in 1933 in New York City, is Higgins Professor of Physics, Harvard University, and Senior Scientist, Smithsonian Astrophysical Observatory, Cambridge, Mass. He is the author of over 100 scientific articles in addition to *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (1972), and *The First Three Minutes: A Modern View of the Origin of the Universe* (1977).

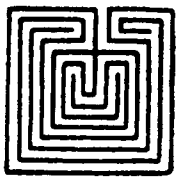
EDWARD O. WILSON, born in 1929 in Birmingham, Ala., is Frank B. Baird, Jr. Professor of Science at Harvard University. He has published *The Insect Societies* (1971), *Sociobiology: The New Synthesis* (1975), and, with Robert H. MacArthur, *The Theory of Island Biogeography* (1967).

GERALD HOLTON, born in 1922, is Mallinckrodt Professor of Physics and Professor of the History of Science at Harvard University. From 1957 to 1963, he was editor of The American Academy of Arts and Sciences and, until 1961, editor of *Daedalus*. Among his book publications are *Thematic Origins of Scientific Thought: Kepler to Einstein* (1973), *Introduction to Concepts and Theories in Physical Science* (1952, 1973), and, as editor, *The Twentieth-Century Sciences: Studies in the Biography of Ideas* (1972), and *Science and Culture* (1965).

PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED

DAEDALUS

INDEX
1977



PRODUCED BY UNZ.ORG
ELECTRONIC REPRODUCTION PROHIBITED